



No. 2001-82

**REDISTRIBUTION AND EDUCATION SUBSIDIES  
ARE SIAMESE TWINS**

By A. Lans Bovenberg and Bas Jacobs

October 2001

ISSN 0924-7815

**Discussion paper**

# Redistribution and Education Subsidies are Siamese Twins\*

A. Lans Bovenberg<sup>†◊</sup> and Bas Jacobs<sup>‡</sup>

October 23, 2001

## Abstract

We develop a model of human capital formation with endogenous labor supply and heterogeneous agents to explore the optimal level of education subsidies along with the optimal progressive schedule of the labor income tax and optimal capital income taxes. Subsidies on education ensure efficiency in human capital accumulation, while taxes on skilled labor help to redistribute income towards the less able. We thus provide a rationale for the widely observed presence of education subsidies. The actually observed tax codes and level of education subsidies suggest that a large part of education subsidies can be justified on these grounds.

Keywords: human capital, education subsidies, progressive taxation, dual income taxation.

JEL codes: H2, H5, I2, J2.

---

\*The authors gratefully thank Michael Reiter, Søren Bo Nielsen, Peter Sørensen and participants of the CEPR/EPRU conference on Dynamic Aspects of Public Expenditures, September 28-30 2001, Copenhagen, Denmark for their comments and suggestions. We benefitted also from comments by participants of seminars held at the CPB Netherlands Bureau of Economic Policy Analysis and the Dutch Central Bank. Bas Jacobs gratefully acknowledges financial support from the NWO Priority Program 'Scholar' funded by the Netherlands Organization for Sciences.

<sup>†</sup>Tilburg University, CentER, Erasmus University, OCfEB, and CEPR. <sup>‡</sup>University of Amsterdam, Tinbergen Institute, Tilburg University, CentER, NWO Priority Program 'Scholar', and CPB Netherlands Bureau for Economic Policy Analysis. <sup>◊</sup>Corresponding author: Department of Economics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: (+31) - 13 - 466 2912. Fax: (+31) - 13 - 466 3042. E-mail: a.l.bovenberg@kub.nl.

# 1 Introduction

Most OECD countries heavily subsidize (higher) education. These education subsidies are typically justified on the basis of the perceived positive external effects of human capital accumulation and capital market imperfections. Positive external effects of higher education, however, are difficult to establish empirically (see, e.g., Heckman and Klenow, 1997; Krueger and Lindahl, 1999; Acemoglu and Angrist, 1999). Moreover, capital market imperfections do not seem to be very important (see, e.g., Shea, 2000; Cameron and Taber, 2000).<sup>1</sup> Why, then, is education subsidized?

We provide a case for education subsidies on the basis of redistributive considerations rather than externalities and capital market imperfections. Although the able benefit more than proportionally from education subsidies, we show that education subsidies play a crucial role so as to alleviate the distortions in human capital accumulation that are induced by redistributive policies. Indeed, our calculations show that these arguments go a long way towards explaining the level of education subsidies in OECD countries.<sup>2</sup>

Our paper explores the interaction between public spending and tax policies by viewing education and tax policies as interdependent instruments aimed at redistribution.<sup>3</sup> The tax literature on the dynamic effects of taxation on the incentives to accumulate human capital<sup>4</sup>, in contrast, has typically abstracted from both public spending and distributional considerations. The

---

<sup>1</sup>In any case, loans rather than subsidies are the most direct way to address liquidity constraints.

<sup>2</sup>An alternative explanation for education subsidies is offered by Boadway et al. (1996) and Andersson and Konrad (2000). They argue that education subsidies are called for if the government cannot commit and engages in excessive redistribution after investments in human capital have been made. We show that education subsidies are part of an optimal redistributive policy mix even if the government can commit to announced policies.

<sup>3</sup>Our paper extends earlier research by Ulph (1977) and Hare and Ulph (1979). They study the problem of optimal non-linear taxation and education expenditures. In both studies, however, the government simply sets the level of education for each agent, so that agents do not choose their levels of learning. Taxation therefore does not distort learning decisions. This contrasts with our analysis in which tax distortions on learning provide the argument for subsidizing education.

<sup>4</sup>See for the literature, e.g., Heckman (1976), Boskin (1977), Kotlikoff and Summers (1979), Eaton and Rosen (1980), Driffil and Rosen (1983), Lord (1989), Nerlove et al. (1993), Pecorino (1993), Trostel (1993), Jones et al. (1993, 1997), Stokey and Rebelo (1995), Bovenberg and Van Ewijk (1997), Nielsen and Sørensen (1997), Milesi-Feretti and Roubini (1998), Hendricks (1999), and Judd (1999), amongst others.

tax literature may therefore have overstated the costs of distortionary taxation in terms of reduced accumulation of human capital and understated the benefits of these taxes in terms of distributional benefits.

We investigate how the availability of education subsidies affects the optimal income tax system. In this connection, we consider both the optimal mix of labor and capital taxation and the optimal progression in marginal tax rates on labor income. We demonstrate that education subsidies make the optimal labor tax more progressive. Moreover, education subsidies eliminate the case for a positive capital income tax as an instrument to stimulate learning. Education subsidies ensure that neither human capital investment nor financial investment are distorted, even though the labor tax is progressive. These subsidies thus eliminate tax distortions on the production side of the economy (see also Diamond and Mirrlees (1971)). Moreover, with separable preferences, the optimal tax on capital income is zero, even allowing for distributional concerns.<sup>5</sup>

In order to study optimal education subsidies along with optimal progressive labor and flat capital taxes, we formulate a two-period life-cycle model of human capital accumulation with heterogeneous agents.<sup>6</sup> Leisure demand is elastic in both periods of the life cycle. In the first period, agents supply unskilled labor and take out education which allows them to supply skilled labor in the second period. Compared to the less able, the more able agents supply more skilled and less unskilled labor and thus concentrate their labor supply more in the second phase of their life cycle. In the first period, they concentrate on acquiring skills rather than supplying unskilled labor. In the second period, the human capital that has been acquired stimulates labor supply because human capital is productive only in work and not in leisure. Indeed, the quantity and quality of second-period labor supply are interdependent.

Our model is closely related to that of Nielsen and Sørensen (1997), who explore the role of progressive taxation as an instrument to offset the excessive learning incentives produced by a positive capital income tax. We extend their model in several directions. First, we allow for heterogeneous households. Accordingly, we do not have to exclude uniform lump-sum taxes

---

<sup>5</sup>This is a familiar result from infinite horizon models (see, e.g., Chamley (1986), Jones et al. (1997), and Judd (1999)) and life-cycle models with separable preferences and homogeneous agents (see Bernheim (1999)).

<sup>6</sup>We focus on intra- rather than intergenerational distribution by employing debt policy to off-set inter-generational inequities.

as a government instrument to make the optimal tax problem interesting. Indeed, it is hard to make sense of (dual) income taxation in the absence of redistributive motives. Moreover, in order to justify progressive taxation in the presence of endogenous leisure demand, we do not have to resort to the assumption that unskilled labor supply is more elastic than skilled labor supply. Instead, the progressive nature of the labor tax system follows immediately from the redistributive motives of the government. More generally, we extend results on the optimal tax structure of capital and labor income taxes in dynamic economies (see, e.g. Atkinson and Sandmo (1980), Nielsen and Sørensen (1997), Jones et al. (1993, 1997), and Judd (1999)) by allowing for heterogeneous agents and distributional concerns.

Second, in contrast to Nielsen and Sørensen (1997), we allow for non-deductible pecuniary outlays (e.g. tuition fees). We show that, in the absence of education subsidies, these non-deductible expenses<sup>7</sup> make a dual income tax (with a progressive labor tax and a positive capital income tax) optimal. The optimal capital income tax is thus positive – even though preferences are additively separable. Given this preference structure, Nielsen and Sørensen (1997) have to rely on the exogenous assumption that the government is committed to a positive tax on capital income to justify a positive capital income tax. Accordingly, if education subsidies are restricted, we provide a stronger underpinning for a dual income tax.

The third extension of Nielsen and Sørensen (1997) is that we allow for education subsidies. We demonstrate that education subsidies eliminate the case for a dual income tax with positive capital taxes if preferences are additively separable. The case for dual income taxation thus depends heavily on restrictions on the use of the instrument of education subsidies.

Our paper is related also to Lommerud (1989), Van Ewijk and Tang (2000) and Dur and Teulings (2001). In Lommerud (1989), the government taxes labor income in order to internalize the negative externalities from status seeking but employs education subsidies to restore incentives to undertake education. In Van Ewijk and Tang (2000), the government employs progressive taxes to punish wage demands of unions and to raise employment, but this discourages learning efforts. Education subsidies allow the government to set progressive labor taxes without distorting human capital accumulation.

---

<sup>7</sup>Non-deductible expenses explain why flat taxes on labor income harm human capital accumulation in several growth models (see, e.g., Pecorino (1993), Stokey and Rebelo (1995), or Milesi-Firetti and Roubini (1998)). In these models, capital goods, which are non-deductible for the labor tax, enter the production function of human capital.

Dur and Teulings (2001) analyze the optimal setting of education subsidies and redistributive income taxes where workers are imperfect substitutes in production. Education subsidies now not only correct tax distortions arising from redistribution, but also exert positive distributional effects by reducing wage differentials between skilled and unskilled workers through an increase in the relative supply of skilled workers.

We extend these papers in a number of important ways. First, we allow for a richer tax code with capital income taxes. Indeed, positive capital income taxes may play a role in encouraging learning, thereby alleviating inadequate learning on account of redistributive taxes on skill. Second, we incorporate non-deductible costs of education, which are observed in many OECD countries. These costs provide an argument for positive capital income taxes if education subsidies are constrained. The introduction of endogenous leisure demand is another extension compared to Van Ewijk and Tang (2000) and Dur and Teulings (2001). We demonstrate that endogenous first-period labor supply is an important factor in explaining: first, the progressive nature of optimal labor taxes; second, the value-added of education subsidies over taxes on unskilled labor as an instrument to alleviate learning distortions due to redistributive policies; and, third, the optimality of a positive capital income tax in the absence of education subsidies.

Many countries have reduced taxes on unskilled labor in recent years in order to encourage unskilled workers to participate in the labor market. This suggests that policy makers are indeed concerned about the adverse employment impact of high taxes on low skilled labor. However, lowering taxes on the unskilled, thereby making the tax system more progressive, harms incentives to acquire skills. We show that the government may want to move the tax burden from the labor market towards the capital market or to introduce education subsidies to offset tax distortions on human capital formation without heavily burdening the employment prospects of the unskilled.

The rest of this paper is structured as follows. Section 2 presents the model. The optimal tax problem is set up in section 3. Section 4 defines distributional characteristics of various tax bases and shows that differences in learning behavior drive heterogeneity in behavior. We study optimal redistributive policies in three steps. First, section 5 investigates optimal labor income taxation if capital income taxes and education subsidies are exogenously given. Subsequently, section 6 turns to optimal dual income taxation by also allowing the government to optimally set capital income taxation. Finally, section 7 introduces education subsidies as an additional instrument.

This section investigates how the availability of the instrument of education subsidies affects optimal tax structures. Section 8 investigates to what extent our model can explain existing education subsidies in several OECD countries. Section 9 concludes. The appendix contains several technical proofs.

## 2 Private behavior

We consider a two-period life-cycle model. Before-tax wage rates and interest rates are exogenously given.<sup>8</sup> A mass of agents with unit measure lives for two periods. In the first period, agents devote their time endowment to supplying unskilled labor, learning, or enjoying leisure. In the second period, the agents spend their time on leisure, which can be interpreted as retirement, and supplying skilled labor. Income can be transferred across the two periods through the accumulation of both financial assets (i.e. financial saving) and human capital (i.e. learning). Capital markets are perfect, so that agents do not suffer from liquidity constraints. Agents are heterogeneous with respect to their ability to learn. The cumulative distribution of the ability to learn  $\alpha$  is denoted by  $F(\alpha)$ .  $F$  has support  $[\underline{\alpha}, \infty)$ . The government knows only the distribution of these abilities, but cannot observe the type of each individual. Accordingly, the government can not levy individual-specific lump-sum taxes, but has to rely on distortionary taxes to redistribute incomes.

In each period of their lives, agents are endowed with one unit of time. They start their lives with one efficiency unit of labor. In the first period, a fraction  $e_\alpha$  of the time endowment is spent on education and a fraction  $h_{1\alpha}$  on leisure, so that a fraction  $l_{1\alpha} = 1 - e_\alpha - h_{1\alpha}$  is left for (unskilled) work. The accumulation of human capital requires not only time, but also commodities. Unit costs of goods per unit of time spent learning amount to  $k$ .<sup>9</sup> Education is subsidized at rate  $s$  per unit of time spent on learning. The before-tax wage rates per efficiency unit of human capital, as well as the price of consumption, are normalized at unity in both periods.<sup>10</sup>

---

<sup>8</sup>The model can thus be viewed either as a partial equilibrium model of a closed economy or a model of a small open economy in which the international capital market fixes the interest rate.

<sup>9</sup>We thus do not allow for substitution between goods and time in the accumulation of human capital. Introducing substitution would merely complicate the analysis without generating additional insights.

<sup>10</sup>Workers earn the same gross wage per unit of human capital, so that workers with different skills are perfect substitutes on the labor market. See Dur and Teulings (2001)

Following Nielsen and Sørensen (1997), we model a labor tax schedule with two brackets. Below an exogenous threshold  $\chi$ , labor income is subject to a tax rate  $t_1$ . Above this threshold, a tax rate  $t_2$  applies. In addition, each agent collects a non-individualized lump-sum transfer, or negative income tax,  $g$ . If  $t_2 > t_1$  the tax system features increasing marginal tax rates on labor income. We assume that first-period labor income (i.e. income from unskilled work) falls only in the first tax bracket. First-period labor income  $l_{1\alpha} = 1 - e_\alpha - h_{1\alpha} < \chi$  is thus taxed at a marginal rate of  $t_1$ . Goods invested in education are not deductible from the labor income tax.<sup>11</sup> Savings  $a_\alpha$  are given by after-tax labor income in the first period, minus consumption and the goods invested in learning (net of subsidies):

$$a_\alpha = (1 - t_1)(1 - e_\alpha - h_{1\alpha}) - c_{1\alpha} - (k - s)e_\alpha. \quad (1)$$

In the second period, human capital is supplied to the labor market in the form of skilled labor.  $\phi(\alpha; e_\alpha)$  is the production function for human capital and measures the number of efficiency units of human capital resulting from learning efforts in the first period  $e_\alpha$ . The production function is given by the following functional form:

$$\phi(\alpha; e_\alpha) \equiv \alpha e_\alpha^\beta. \quad (2)$$

This production function features positive but diminishing returns to time invested in education (i.e.  $\phi_e > 0$  and  $\phi_{ee} < 0$ ). Furthermore, ability facilitates learning (i.e.  $\phi_\alpha > 0$ ), while high ability agents are relatively more efficient in the production of human capital (i.e.  $\phi_{e\alpha} > 0$ ). In the second period, a fraction  $h_{2\alpha}$  of the time endowment is devoted to leisure, while the rest (i.e.  $l_{2\alpha} = 1 - h_{2\alpha}$ ) is spent working. Before-tax labor income (from skilled work)  $l_{2\alpha}\phi(\alpha; e_\alpha)$  exceeds the tax threshold  $\chi$ , so that it is subject to a marginal tax rate of  $t_2$ , i.e.  $l_{2\alpha}\phi(\alpha; e_\alpha) > \chi \forall \alpha$ .<sup>12</sup> Income from skilled work can thus be taxed at a different marginal rate than income from unskilled work (if  $t_1 \neq t_2$ ).

Income derived from accumulation of financial assets is  $(1 + r(1 - \tau))a_\alpha$ , where  $r$  stands for the exogenous real interest rate and  $\tau$  denotes the tax

---

for optimal education policies and redistribution in a general equilibrium model in which various skills are imperfect substitutes in demand.

<sup>11</sup>Tax deductible expenses can be modelled by setting  $s = t_1 k$ .

<sup>12</sup>This assumption puts a lower bound on ability  $\alpha$  to avoid bunching induced by the kink in the private budget constraint if the second-period income tax  $t_2$  exceeds the first-period income tax  $t_1$  (i.e.  $t_2 > t_1$ ). See below.



rate on capital income. All income from human and financial sources is spent on consumption  $c_2$ . Second-period consumption is untaxed.<sup>13</sup> Hence, the second-period budget constraint amounts to:

$$c_{2\alpha} = (1 - t_2)l_{2\alpha}\phi(\alpha; e_\alpha) + Ra_\alpha + (t_2 - t_1)\chi + g, \quad (3)$$

where  $R \equiv 1 + (1 - \tau)r$ .

Households derive utility from consumption and leisure according to the following quasi-linear utility function:

$$\begin{aligned} u(c_{1\alpha}, h_{1\alpha}, c_{2\alpha}, h_{2\alpha}) = & (1 + \rho) \left( \frac{c_{1\alpha}^{1-1/\varepsilon_c}}{1 - 1/\varepsilon_c} - \frac{(1 - h_{1\alpha})^{1+1/\varepsilon_1}}{1 + 1/\varepsilon_1} \right) \\ & + c_{2\alpha} - \frac{(1 - h_{2\alpha})^{1+1/\varepsilon_2}}{1 + 1/\varepsilon_2}, \end{aligned} \quad (4)$$

where  $\rho$  represents the pure rate of time-preference and  $\varepsilon_c, \varepsilon_1, \varepsilon_2 > 0$  are parameters. The specific utility function implies that income effects in labor supply and first-period consumption are absent.<sup>14</sup> Furthermore, our additive structure implies that various cross-substitution effects drop out. This particular specification of utility does not affect our main results but is used for ease of exposition. Indeed, the appendix shows that the result that education subsidies eliminate learning distortions (derived in section 7) does not depend on the specific utility function (4), but holds also with a general utility function  $u(c_{1\alpha}, h_{1\alpha}, c_{2\alpha}, h_{2\alpha})$  that allows for cross-substitution and income effects.<sup>15</sup>

Agents maximize utility by choosing  $e_\alpha, c_{1\alpha}, l_{1\alpha}, c_{2\alpha}, l_{2\alpha}$  and  $a_\alpha$ , taking the instruments of the government as given. Indeed, the government is assumed to set policy before agents determine their behavior.<sup>16</sup> The first-order

---

<sup>13</sup>The tax system can be normalized in a different way, but this does not affect the optimal allocation.

<sup>14</sup>When studying optimal non-linear income taxation in a static model of labor supply, Diamond (1998) adopts a similar approach by assuming that income effects are absent in labor supply.

<sup>15</sup>In contrast to Heckman (1976), this specification of the utility function implies that human capital is not productive in leisure time. Human capital is an investment rather than a consumption good.

<sup>16</sup>In view of its distributional preferences, the government faces an incentive to renege on its promises after the private sector has accumulated human and financial capital. We thus have to assume that the government has access to a commitment technology (e.g. due to reputational considerations). For the case for education subsidies in case the government cannot commit, see Boadway et al. (1996) and Anersson and Konrad (2000).

condition for optimal learning of the private household amounts to:

$$(1 - t_2)l_{2\alpha}\phi_e(.) = (1 + r(1 - \tau))(1 - t_1 + k - s). \quad (5)$$

The condition states that marginal benefits of an hour's learning, i.e. the marginal increments in after-tax second-period income (see the left-hand side of (5)), should be equal to the marginal costs of learning (i.e. the second-period value of forgone earnings plus out-of-pocket education expenses (net of subsidies) in the first period, see the right-hand side of (5)). The positive cross derivative  $\phi_{e\alpha}$  implies that high-ability agents choose to learn more than low-ability households do.

The first-order condition for learning can also be interpreted as an arbitrage condition between the returns on investing a unit of time in learning (i.e. the left-hand side of (5)) and the returns on investing a unit of time in financial capital (by working and investing the rewards – including the saved out-of-pocket expenditures on education (net of subsidies)  $k - s$  – in the capital market, i.e. the right-hand side of (5)).

The first-order conditions for consumption and leisure yield the following solutions for first-period consumption and leisure in both periods:

$$c_1 = \left( \frac{1 + \rho}{1 + r(1 - \tau)} \right)^{\varepsilon_c}, \quad (6)$$

$$h_1 = 1 - \left( \frac{(1 + r(1 - \tau))(1 - t_1)}{1 + \rho} \right)^{\varepsilon_1}, \quad (7)$$

$$h_{2\alpha} = 1 - ((1 - t_2)\phi(.))^{\varepsilon_2}. \quad (8)$$

Demand for first-period consumption increases with the capital income tax  $\tau$ , as the intertemporal substitution effect of the lower after-tax interest rate dominates the absent income effect. Similar intertemporal substitution effects cause a higher capital income tax to boost the demand for first-period leisure. This demand is also increased by a substitution effect on account of a higher marginal tax rate on first-period labor income  $t_1$ . In the absence of income effects, all agents demand the same amount of consumption and leisure in the first period, since everybody faces the same net interest rate and first-period wages. Second-period leisure, in contrast, is lowest for high-ability agents, who benefit from higher second-period wage rates  $(1 - t_2)\phi(.)$  on account of their higher learning efforts in the first period.

Solving (5) and (8) for learning  $e_\alpha$  and second-period labor-supply  $l_{2\alpha} = 1 - h_{2\alpha}$  by using the functional form (2) for the production function for human capital, we arrive at closed-form solutions for  $e_\alpha$  and  $l_{2\alpha}$ :<sup>17</sup>

$$e_\alpha = \left( \frac{\alpha^{1+\varepsilon_2} \beta (1-t_2)^{1+\varepsilon_2}}{R(1-t_1+k-s)} \right)^{1/(1-\beta(1+\varepsilon_2))}, \quad (9)$$

$$l_{2\alpha} = (1-t_2)^{\varepsilon_2} (\alpha e_\alpha^\beta)^{\varepsilon_2}. \quad (10)$$

Second-period labor income  $l_{2\alpha}\phi(\cdot)$  rises with ability. Indeed, this income from skilled labor is proportional to learning:<sup>18</sup>

$$l_{2\alpha}\phi(\cdot) = \mu e_\alpha, \quad (11)$$

where the proportionality factor  $\mu \equiv \frac{R(1-t_1+k-s)}{\beta(1-t_2)}$  does not depend on skill  $\alpha$ . These solutions imply also that the elasticities of learning and second-period labor supply with respect to the policy instruments are the same for all agents. In contrast to second-period labor supply (i.e. skilled labor supply), first-period labor supply  $l_{1\alpha} = 1 - e_\alpha - h_1$  (i.e. unskilled labor supply) declines with ability – as all agents demand the same leisure  $h_1$  but high ability agents spend more time on learning  $e_\alpha$ . Whereas able agents tend to concentrate labor supply at the end of their lives, less able agents work relatively more in the beginning of their lives. The proportionality factor between after-tax labor income (from unskilled labor) in the first period,  $(1-t_1)l_{1\alpha}$ , and learning  $e_\alpha$  is given by  $-(1-t_1)$  and thus also independent of skill.

The second-order condition for a maximum requires (see appendix):

$$\beta(1+\varepsilon_2) < 1, \quad (12)$$

so that learning and second-period labor supply decline with the second-period tax rate. The second-order condition guarantees an interior solution for learning by ensuring that the returns to learning decline if more time is devoted to learning. The positive feedback effects between human capital and

---

<sup>17</sup>We assume that the parameters are such that  $e_\alpha < 1 - h_1 = \left( \frac{(1+r(1-\tau))(1-t_1)}{1+\rho} \right)^{\varepsilon_1}$ .

<sup>18</sup>This linear relationship follows immediately from (5) and the Cobb-Douglas learning function (2) and thus does not depend on the specification of utility (see also the appendix). (11) together with (9) implies that the no-bunching constraint  $l_{2\alpha}\phi(\cdot) > \chi$  can be written as  $\mu \left( \frac{\alpha^{1+\varepsilon_2} \beta (1-t_2)^{1+\varepsilon_2}}{R(1-t_1+k-s)} \right)^{1/(1-\beta(1+\varepsilon_2))} > \chi$ . This constraint implies a lower bound on  $\alpha$ .

second-period labor supply imply that decreasing returns in the production function of human capital (i.e.  $\beta < 1$ ) are not sufficient for this second-order condition to be met. In particular, more learning raises second period labor supply (if  $\varepsilon_2 > 0$ ), which in turn makes learning more attractive. This positive feedback effect, which depends on the wage elasticity of second-period labor supply  $\varepsilon_2$ , should be offset by sufficiently strong decreasing returns in the production function of human capital to prevent corner solutions.

The positive feedback effects between learning and labor supply makes both the learning decision and the labor supply response more elastic. In particular, the interaction between the quality and quantity of second-period labor supply raises the absolute value of the after-tax wage elasticity of second-period labor supply from  $\varepsilon_2$  in a model without learning (i.e.  $\beta = 0$ ) to  $\frac{\varepsilon_2}{1-\beta(1+\varepsilon_2)}$ .<sup>19</sup> Similarly, endogenous second-period labor supply makes learning more sensitive to the second-period wage rate (the wage elasticity of learning is  $-\frac{(1+\varepsilon_2)}{1-\beta(1+\varepsilon_2)}$  compared to an elasticity of only  $\frac{-1}{1-\beta}$  in a model with exogenous labor supply (i.e. in which  $\varepsilon_2 = 0$ )).

Second-period leisure can be interpreted as retirement. The solutions thus reveal an important interaction between human capital accumulation and retirement behavior. On the one hand, early retirement discourages the accumulation of human capital because it reduces the returns on learning. On the other hand, a lack of schooling encourages early retirement on account of low labor productivity. Indeed, the quantity and quality of labor supply are closely related.

### 3 Government

The government collects taxes to finance exogenously given expenditures in the second period,  $\Lambda$ , the education subsidy  $s$ , and the uniform lump-sum transfer  $g$ . The government budget constraint therefore reads as:

$$\begin{aligned} & \int_{\underline{\alpha}}^{\infty} [t_1(1+r)(1-e_{\alpha}-h_1) + t_2(1-h_{2\alpha})\phi(\alpha; e_{\alpha})] dF(\alpha) \\ & + \int_{\underline{\alpha}}^{\infty} \tau r a_{\alpha} dF(\alpha) = \int_{\underline{\alpha}}^{\infty} [(1+r)se_{\alpha} + (t_2 - t_1)\chi + g + \Lambda] dF(\alpha). \end{aligned} \quad (13)$$

---

<sup>19</sup>Hence, if in the absence of learning the wage elasticities of labor supply are the same in both periods (ie.  $\varepsilon \equiv \varepsilon_1 = \varepsilon_2$ ), endogenous learning increases the wage elasticity of labor supply in the second period above the corresponding elasticity in the first period.

Employing the definition of private savings (1), we can rewrite the government budget constraint in terms of the bases of the labor taxes:

$$\int_{\underline{\alpha}}^{\infty} [t_1 (R(1 - e_{\alpha} - h_1) + \chi) + t_2 ((1 - h_{2\alpha})\phi(\alpha; e_{\alpha}) - \chi)] dF(\alpha) \quad (14)$$

$$+ \int_{\underline{\alpha}}^{\infty} \tau r(1 - c_1 - (1 + k)e_{\alpha} - h_1) dF(\alpha) = \int_{\underline{\alpha}}^{\infty} [Rse_{\alpha} + g + \Lambda] dF(\alpha).$$

For each generation, the government's budget is fully funded. In contrast to an approach that maximizes steady-state social utility subject to a steady-state government budget constraint (see e.g. King (1980) and Sandmo (1985)), this procedure does not ignore the welfare effects of generations living through the transition to a new tax system. Our specification of the government budget constraint implies that the government can not raise the welfare of steady-state generations by transferring resources away from generations living through the transition. We thus not only clearly isolate efficiency impacts from effects on the intergenerational distribution of resources, but also model grandfathering schemes in actual tax reforms protecting agents that have not been able to anticipate the change in the tax rules (see also Nielsen and Sørensen (1997)).<sup>20</sup>

The government maximizes a social welfare function  $\Gamma$ :

$$\Gamma = \int_{\underline{\alpha}}^{\infty} \Psi(v_{\alpha}) dF(\alpha), \quad \Psi' > 0, \quad \Psi'' \leq 0, \quad (15)$$

where  $v_{\alpha}$  stands for the indirect utility function of an agent with skill  $\alpha$ . The concavity of  $\Psi$  reflects the strength of the redistributive preferences of the government. If  $\Psi$  is linear, the government maximizes a utilitarian social welfare function. Together with the quasi-linear private preferences (4), this implies that the government features no distributional concerns.

## 4 Distributional characteristics

The interpretation of the optimal policy rules is facilitated by defining distributional characteristics of the various tax (and subsidy) bases. The distributional characteristic of a tax base is given by the negative normalized

---

<sup>20</sup>Indeed, the government is assumed to have access to enough instruments to insulate the current generations from the reform. Debt policy suffices for this purpose in our two-period life cycle model. However, if generations would live for more than two periods, the government would have to levy age-specific taxes to be able to protect current generations.

covariance between the welfare weight the government attaches to life-time income of a particular skill  $b_\alpha$  (which is non-increasing with the skill level  $\alpha$ ) and the marginal contribution of agent  $\alpha$  to the tax base  $y_\alpha$  (see, e.g., Atkinson and Stiglitz (1980)):

$$\xi_y \equiv - \frac{\left( \int_{\underline{\alpha}}^{\infty} y_\alpha b_\alpha dF(\alpha) - \int_{\underline{\alpha}}^{\infty} y_\alpha dF(\alpha) \int_{\underline{\alpha}}^{\infty} b_\alpha dF(\alpha) \right)}{\int_{\underline{\alpha}}^{\infty} y_\alpha dF(\alpha) \int_{\underline{\alpha}}^{\infty} b_\alpha dF(\alpha)}. \quad (16)$$

A positive distributional characteristic thus implies that the tax base is larger for high skills (which feature low welfare weights) than for low skills, so that taxing this base generates positive distributional benefits. The magnitude of a distributional characteristic depends both on the correlation between skills and the tax base and the strength of the redistributive preferences as reflected in the negative correlation between skills and the welfare weights.<sup>21</sup> Indeed, a distributional characteristic of zero may indicate either that the government is not interested in redistribution (so that all skills feature the same welfare weight) or that the marginal contribution to the tax base is the same for all skills.

The distributional characteristics are closely related because the tax (and subsidy) bases are all linearly related to learning  $e_a$ . In particular, first-period labor income (which is subject to  $t_1$ ) is given by  $1 - h_1 - e_\alpha$ , second-period labor income (which is subject to  $t_2$ ) by  $\frac{R(1-t_1+k-s)}{\beta(1-t_2)} e_\alpha$  (see (11)), and savings (which is subject to  $\tau$ ) by  $(1 - t_1)(1 - h_1) - c_1 - (1 - t_1 + k - s)e_\alpha$ . Hence, in the second period, more able agents enjoy higher labor income than less able agents do because they earn higher wages and work more hours. In the first period, in contrast, the least able agents earn the highest labor incomes, and therefore save most, because they spend less time learning (and all agents enjoy the same amount of consumption  $c_1$  and leisure  $h_1$ ). Using these relationships, we arrive at the following lemma describing the relationships between the various distributional characteristics. Here, the subscripts 1, 2,  $e$  and  $a$  represent first-period labor income (i.e. income from

---

<sup>21</sup>The strength of this negative correlation depends not only on the concavity of the function  $\Psi$ , but in general also on inequality in life-time incomes. In particular, the government attaches a higher priority to redistributing incomes if life-time incomes become more unequal, since marginal utility of income declines with income. However, in the current set-up, all distributional motives enter through  $\Psi$  since marginal utility is constant and equal to unity for all agents because of the quasi-linear utility specification.

unskilled labor), second-period labor income (i.e. income from skilled labor), learning and savings, respectively.

**Lemma**  $\xi \equiv \xi_2 = \xi_e = -\xi_1 \frac{\int_{\underline{\alpha}}^{\infty} (1-e_{\alpha}-h_1)dF(\alpha)}{\int_{\underline{\alpha}}^{\infty} e_{\alpha}dF(\alpha)} = -\frac{\xi_a}{(1-t_1+k-s)} \frac{\int_{\underline{\alpha}}^{\infty} a_{\alpha}dF(\alpha)}{\int_{\underline{\alpha}}^{\infty} e_{\alpha}dF(\alpha)} > 0$ .

Proof: see appendix.

The distributional characteristics of learning and second-period labor income (i.e.  $\xi_2$  and  $\xi_e$  respectively) are positive because the richest agents (i.e. the most able) learn more and earn more second-period labor income. Since these agents save less and earn less first-period labor income, the corresponding characteristics for financial savings<sup>22</sup> and first-period labor income (i.e.  $\xi_a$  and  $\xi_1$  respectively) are negative.

## 5 Optimal labor income taxation

The Lagrangean for maximization of social welfare is given by:

$$\begin{aligned} \mathcal{L} = & \int_{\underline{\alpha}}^{\infty} \Psi(v_{\alpha})dF(\alpha) \\ & + \eta \int_{\underline{\alpha}}^{\infty} [t_1 (R(1 - e_{\alpha} - h_1) + \chi) + t_2 ((1 - h_{2\alpha})\phi(\alpha; e_{\alpha}) - \chi)] dF(\alpha) \\ & + \eta \int_{\underline{\alpha}}^{\infty} \tau r(1 - c_1 - (1 + k)e_{\alpha} - h_{1\alpha})dF(\alpha) \\ & - \eta \int_{\underline{\alpha}}^{\infty} [Rse_{\alpha} + g + \Lambda] dF(\alpha), \end{aligned}$$

where  $\eta$  represents the Lagrange multiplier associated with the government budget constraint. This section explores the case in which the government can freely set the parameters of the labor-income tax (i.e.  $g$ ,  $t_1$  and  $t_2$ ) for a given capital income tax,  $\tau$ , and a given learning subsidy  $s$ .

---

<sup>22</sup>This assumes that economy-wide saving is positive. If saving is negative, the distributional characteristic is positive – even though the negative covariance is negative because the denominator of the normalized covariance is negative.

## 5.1 Optimal lump-sum transfer

The first-order condition for the optimal lump-sum element  $g$  amounts to:<sup>23</sup>

$$\int_{\underline{\alpha}}^{\infty} b_{\alpha} dF = 1. \quad (17)$$

where  $b_{\alpha} \equiv \Psi'/\eta$  stands for the welfare weight of marginal life-time income of each of the agents normalized by the marginal value of government revenue. In deriving (17), we used  $\partial v_{\alpha}/\partial g = 1$  (i.e. Roy's identity) and the fact that income effects are absent in first-period consumption, leisure demands and learning so that these variables do not depend on the lump-sum transfer (i.e.  $\partial c_1/\partial g = \partial h_1/\partial g = \partial h_2/\partial g = \partial e/\partial g = 0$ , see also (6), (7), (8), (9)). Expression (17) shows that in the optimum the benefits of higher uniform lump-sum transfers (averaged over all agents, see the left-hand side of (17)) should equal the costs in terms of additional government spending (i.e. the right-hand side of (17)).<sup>24</sup>

## 5.2 Optimal tax on unskilled labor

The first-order condition for the optimal labor tax in the first period (i.e. the marginal tax rate on unskilled labor),  $t_1$ , amounts to (using  $\frac{\partial c_1}{\partial t_1} = 0$  (see (6) and  $\partial v_{\alpha}/\partial t_1 = -R(1 - e_{\alpha} - h_1) - \chi$  (Roy's identity)):

$$\begin{aligned} & \int_{\underline{\alpha}}^{\infty} (b_{\alpha} - 1) \left( (1 - e_{\alpha} - h_1) + \frac{\chi}{R} \right) dF \\ &= \frac{\Delta}{R} \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{\partial t_1} dF - \left( t_1 + \frac{\tau r}{R} \right) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1}{\partial t_1} dF + \int_{\underline{\alpha}}^{\infty} \left( \frac{t_2 \phi(\alpha; e_{\alpha})}{R} \frac{\partial l_2}{\partial t_1} \right) dF, \end{aligned}$$

where  $(t_1 + \tau r/R)$  stands for the subsidy wedge on first-period leisure and  $\Delta \equiv \frac{(1-t_1+k-s)R}{1-t_2} - (1+r)(1+k)$  represents the tax wedge on learning (i.e. the first-order effect of learning  $e_{\alpha}$  on the government budget constraint). Substituting (17), we arrive at:

$$\begin{aligned} & \xi_e \int_{\underline{\alpha}}^{\infty} e_{\alpha} dF - \frac{\Delta}{R} \frac{\varepsilon_{et_1}}{(1-t_1)} \int_{\underline{\alpha}}^{\infty} e_{\alpha} dF \\ &= - \left( t_1 + \frac{\tau r}{R} \right) \frac{\varepsilon_{(1-h_1)t_1}}{(1-t_1)} (1-h_1) + \frac{t_2}{R} \frac{\varepsilon_{l_2 t_1}}{(1-t_1)} \int_{\underline{\alpha}}^{\infty} l_{2\alpha} \phi(\alpha; e_{\alpha}) dF, \end{aligned}$$

<sup>23</sup>The rest of this paper drops the index  $\alpha$  whenever convenient.

<sup>24</sup>In the absence of income effects, the marginal costs of public funds, which can be defined by  $1/[\int b_{\alpha} dF]$ , thus equals unity.



where we have used the fact that the following elasticities do not depend on skill:  $\varepsilon_{et_1} \equiv \frac{\partial e}{\partial t_1} \frac{(1-t_1)}{e} = \frac{(1-t_1)}{(1-\beta(1+\varepsilon_2))(1-t_1+k-s)}$  (the last equality follows from (9)),  $\varepsilon_{(1-h_1)t_1} \equiv \frac{\partial h_1}{\partial t_1} \frac{(1-t_1)}{(1-h_1)} = \varepsilon_1$  (the last equality follows from (7)), and  $\varepsilon_{l_2 t_1} \equiv \frac{\partial l_2}{\partial t_1} \frac{(1-t_1)}{l_2} = \frac{\beta \varepsilon_2 (1-t_1)}{(1-\beta(1+\varepsilon_2))(1-t_1+k-s)}$  (the last equality follows from (8)). Employing (11) to eliminate economy-wide skilled labor income, we arrive at

$$\xi_e + \left(t_1 + \frac{\tau r}{R}\right) \frac{E_{1t_1}}{(1-t_1)} = \frac{\Delta}{R} \frac{\varepsilon_{et_1}}{(1-t_1)} + \frac{t_2}{R} \frac{\mu \varepsilon_{l_2 t_1}}{(1-t_1)}, \quad (18)$$

where  $E_{1t_1} \equiv \varepsilon_{(1-h_1)t_1} \frac{(1-h_1)}{\int_{\alpha}^{\infty} e_{\alpha} dF} = \varepsilon_1 \frac{(1-h_1)}{\int_{\alpha}^{\infty} e_{\alpha} dF} > 0$ . Intuitively, raising the tax on unskilled labor,  $t_1$ , imposes distributional losses (the first term on the left-hand side of (18)) and yields a first-order welfare loss by worsening the distortions in first-period leisure demand (assuming that  $(t_1 + \tau r/R) > 0$ ; see the second term on the left-hand side of (18)). At the optimal tax rate, the distributional and efficiency costs should be equal to the efficiency benefits of raising the tax rate on unskilled labor. These benefits consist of first-order welfare gains due to more learning (assuming that  $\Delta > 0$ ; see the first term on the right-hand side of (18)) and to higher skilled labor supply (assuming that  $t_2 > 0$ ; see the second term at the right-hand side of (18)). Intuitively, skilled and unskilled labor supply are substitutes. Hence, by acting as a subsidy on skilled labor, a tax on unskilled labor alleviates distortions in the market for skilled labor by boosting learning and the supply of skilled labor.

In the absence of other taxes (i.e.  $t_2 = \tau = 0$ ) and with tax deductible education expenses (i.e.  $s = t_1 k$ ), unskilled labor is subsidized:

$$\frac{t_1}{1-t_1} = \frac{-\xi}{[E_{1t_1} + \frac{1}{1-\beta(1+\varepsilon_2)}]}. \quad (19)$$

The reason for this subsidy on unskilled labor is that it helps to alleviate the inequities in lifetime incomes. The subsidy rises with the distributional characteristic  $\xi$  and falls with the elasticities in labor supply  $E_{1t_1}$  and learning  $\frac{1}{1-\beta(1+\varepsilon_2)}$ .

### 5.3 Optimal tax on skilled labor

The first-order condition for the second-period labor tax rate (i.e. the marginal tax rate on skilled labor),  $t_2$ , is given by (using  $\frac{\partial c_1}{\partial t_2} = \frac{\partial h_1}{\partial t_2} = 0$  (see (6) and

(7), respectively) and Roy's identity  $\partial v_\alpha / \partial t_2 = -l_{2\alpha} \phi(\alpha; e_\alpha) + \chi$ ):

$$\begin{aligned} & \int_{\underline{\alpha}}^{\infty} (-b_\alpha + 1) (l_{2\alpha} \phi(\alpha; e_\alpha) - \chi) dF + \Delta \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{\partial t_2} dF \\ &= - \int_{\underline{\alpha}}^{\infty} \left( t_2 \phi(\alpha; e_\alpha) \frac{\partial l_2}{\partial t_2} \right) dF. \end{aligned}$$

Substituting (17), we arrive at:

$$\begin{aligned} & \xi_2 \int_{\underline{\alpha}}^{\infty} l_{2\alpha} \phi(\alpha; e_\alpha) dF - \Delta \frac{\varepsilon_{et_2}}{(1-t_2)} \int_{\underline{\alpha}}^{\infty} e_\alpha dF \\ &= t_2 \frac{\varepsilon_{l_2 t_2}}{(1-t_2)} \int_{\underline{\alpha}}^{\infty} l_{2\alpha} \phi(\alpha; e_\alpha) dF. \end{aligned}$$

where we have used the fact that  $\varepsilon_{et_2} \equiv -\frac{\partial e}{\partial t_2} \frac{(1-t_2)}{e} = \frac{1+\varepsilon_2}{1-\beta(1+\varepsilon_2)}$  (the last equality follows from (9)) and  $\varepsilon_{l_2 t_2} \equiv -\frac{\partial l_2}{\partial t_2} \frac{(1-t_2)}{l_2} = \frac{\varepsilon_2}{1-\beta(1+\varepsilon_2)}$  (the last equality follows from (9) and (10)) do not depend on skill  $\alpha$ . Employing (11) to eliminate economy-wide learning  $\int_{\underline{\alpha}}^{\infty} e_\alpha dF$ , we find:

$$\xi_2 = \frac{\Delta}{\mu} \frac{\varepsilon_{et_2}}{(1-t_2)} + t_2 \frac{\varepsilon_{l_2 t_2}}{(1-t_2)}. \quad (20)$$

The distributional benefits of a higher tax rate (i.e. the left-hand side of (20)) should correspond to the additional first-order welfare losses as a result of the high tax rate (i.e. the right-hand side of (20)). These welfare losses are the sum of the impact on the learning distortion  $\Delta$  ( $\Delta > 0$  if additional learning yields a first-order gain in welfare) and the distortion in second-period labor supply  $t_2$ . In case other tax and subsidies are absent (i.e.  $t_1 = \tau = s = 0$ , so that  $\frac{\Delta}{\mu} = \beta t_2$ ), we can employ (20) to solve for  $t_2$  (by using the definitions of the elasticities  $\varepsilon_{et_2} \equiv \frac{1+\varepsilon_2}{1-\beta(1+\varepsilon_2)}$  and  $\varepsilon_{l_2 t_2} \equiv \frac{\varepsilon_2}{1-\beta(1+\varepsilon_2)}$ ):

$$\frac{t_2}{1-t_2} = \frac{\xi}{\bar{\varepsilon}}, \quad (21)$$

where  $\bar{\varepsilon} \equiv \frac{\varepsilon_2 + \beta(1+\varepsilon_2)}{1-\beta(1+\varepsilon_2)}$  denotes the combined elasticity of learning and skilled labor supply with respect to the reward of supplying skilled labor. In accordance with the standard Ramsey intuition, the optimal tax rate on skilled labor declines with this combined elasticity  $\bar{\varepsilon}$  and rises with the effectiveness of the tax on skilled labor in alleviating life-time income inequality (as captured by the distributional characteristic  $\xi$ ).

## 5.4 Optimal labor tax schedule

Substitution of (18) into (20) to eliminate the distributional characteristics  $\xi_2 = \xi_e$  (and using  $\frac{\mu\varepsilon_{l_2}t_1}{R(1-t_1)} = \frac{\varepsilon_{l_2}t_2}{(1-t_2)}$  and  $\frac{\varepsilon_{et_2}}{\mu(1-t_2)} = \frac{\varepsilon_{et_1}}{(1-t_1)R}\beta(1+\varepsilon_2)$ ) yields:

$$\left(\frac{t_1 + \frac{\tau r}{R}}{1 - t_1}\right) E_{1t_1} = \frac{\Delta}{R(1 - t_1 + k - s)}. \quad (22)$$

Substituting this back in (20) to eliminate the learning wedge  $\Delta$ , we arrive at:

$$\xi_2 = \frac{\beta(1 + \varepsilon_2)}{1 - \beta(1 + \varepsilon_2)} \left(t_1 + \frac{\tau r}{R}\right) E_{1t_1} + t_2 \frac{\varepsilon_{l_2}t_2}{(1 - t_2)}. \quad (23)$$

Together with (22), (23) determines the optimal tax structure. In the specific case that education expenses are tax deductible (i.e.  $s = t_1k$ ), we can solve for  $t_1$ ,  $t_2$ , and the learning wedge  $\Delta/(R(1 - t_1)(1 + k))$  (see the appendix):

$$\frac{t_2}{1 - t_2} = \frac{\xi}{\varepsilon^*}, \quad (24)$$

$$\frac{t_1 + \frac{\tau r}{R}}{1 - t_1} = \frac{t_2}{1 - t_2} \frac{1}{(1 + E_{1t_1}^*)} = \frac{\xi}{(1 + E_{1t_1}^*)\varepsilon^*}, \quad (25)$$

$$\frac{\Delta}{R(1 - t_1)(1 + k)} = \frac{E_{1t_1}^*\xi}{(1 + E_{1t_1}^*)\varepsilon^*}, \quad (26)$$

where  $\varepsilon^* \equiv \frac{\varepsilon_2 + \beta(1 + \varepsilon_2)E_{1t_1}/(1 + E_{1t_1})}{1 - \beta(1 + \varepsilon_2)}$  and  $E_{1t_1}^* = E_{1t_1}/(1 + k)$ .

In interpreting the optimal labor tax structure, we first turn to the case in which first-period leisure demand is fixed (i.e.  $E_{1t_1}^* = 0$ ) and the capital income tax is absent (i.e.  $\tau = 0$ ).<sup>25</sup> In this case, a flat tax is optimal (i.e.  $t_1 = t_2$ ). Such a flat tax acts like a pure cash-flow tax on human capital investment. The inframarginal returns on skill are taxed without distorting the incentives to accumulate human capital. Indeed, in the presence of such a pure profit tax on skill, learning is not distorted (i.e.  $\Delta = 0$ ). As a direct consequence, *ceteris paribus* the distributional characteristic  $\xi$ , the tax rate on skilled labor can be higher than in the absence of an instrument to offset learning distortions (compare (21) and (24) and note that  $\bar{\varepsilon} > \varepsilon^* \equiv \frac{\varepsilon_2}{1 - \beta(1 + \varepsilon_2)}$  if learning is endogenous (i.e.  $\beta > 0$ )).

---

<sup>25</sup>The interpretation of the optimal labor tax schedule is equivalent when education expenses are not deductible, even though there is not a closed form solution.

Unskilled labor is taxed (i.e.  $t_1 > 0$ ) even though it is subsidized in the absence of other taxes (compare (19) and (25)) Indeed, unskilled labor is taxed although this hurts equity. The reason is that the tax rate on skilled labor  $t_2$  is a more efficient instrument to even out the lifetime income distribution, while the tax rate on unskilled labor is most efficient at alleviating the learning distortions induced by the redistributive tax on skilled labor. In line with the targeting principle, therefore, the skilled tax is aimed at correcting the income distribution, while the unskilled tax deals with offsetting the learning distortion.

The presence of a positive capital income tax (i.e.  $\tau > 0$ ) affects neither the optimal tax on skilled labor  $t_2$  nor the result of a zero learning distortion  $\frac{\Delta}{R(1-t_1)(1+k)}$ . However, it reduces the optimal tax on unskilled labor  $t_1$  below the tax rate on skilled labor, so that marginal taxes rise with income (i.e.  $t_2 > t_1$ ). Intuitively, the capital income tax favors human capital investment over other types of saving. A progressive labor income tax<sup>26</sup> that taxes skilled labor relatively heavily offsets this distortion in favor of learning. In the context of a model with homogeneous households, Nielsen and Sørensen (1997) employ this argument to argue in favor of a dual income tax in which labor income is taxed at progressive marginal rates if the government is committed to taxing capital income at positive rates. We show that this argument holds also in a setting with heterogeneous households (implying vertical equity considerations) and relatively elastic skilled labor supply.

Our results both strengthen and weaken the results of Nielsen and Sørensen (1997). We weaken their results by showing that progressive taxation is called for only if households are heterogeneous and the government features redistributive preferences (so that the distributional characteristic  $\xi$  is positive).<sup>27</sup> With homogeneous households, the government does not have to employ the distortionary labor income tax to change the income distribution, but can rely only on the non-distortionary lump-sum tax (i.e. the instrument  $g$ ) to

---

<sup>26</sup>A progressive tax is often defined as a tax under which *average* tax rates rise with taxable income. We, in contrast, use the term to mean that *marginal* tax rates increase with taxable income.

<sup>27</sup>Also non deductible education expenses (i.e.  $k > 0$ ,  $s = 0$ ) weaken the case for progressive taxation as an instrument to offset the excessive learning incentives on account of the capital income tax. The reason is that these non-deductable expenses already help to reduce the incentives to accumulate human capital. Education subsidies ( $s > t_1 k$ ), in contrast, encourage agents to train themselves, thereby strengthening the case for progressive labor taxation.

finance all its expenditures.

Nielsen and Sørensen (1997) can establish their main result that labor taxation should be progressive only if unskilled labor supply is more elastic than skilled labor supply (in terms of the parameters of our model this implies that  $\varepsilon_1$  is large compared to  $\varepsilon_2$ ). We strengthen their result by demonstrating that the result holds true even if the elasticity of unskilled labor supply  $\varepsilon_1$  is small compared to the elasticity of skilled labor supply  $\varepsilon_2$ . In the context of the model developed by Nielsen and Sørensen (1997), inelastic unskilled labor supply would provide an argument for levying a relatively heavy tax on unskilled labor (i.e. first-period labor supply). In our model, in contrast, the tax rate on unskilled labor does not exceed the tax rate on skilled labor if unskilled labor supply is relatively less elastic because not only efficiency but also distributional considerations determine optimal tax policy. In particular, whereas a high tax on unskilled labor imposes less distortions on labor supply, it also widens inequities in life-time incomes. Our model thus provides stronger arguments for progressive taxation.

Education subsidies (i.e.  $s > 0$ ) have similar effects as the capital income tax on the optimal progression of the labor income tax. In particular, with exogenous first-period leisure demand, learning is not distorted and the optimal tax system is progressive (i.e.  $t_1 = t_2 - s$  if  $\tau = k = 0$ , see (22) and use the definition of  $\Delta \equiv \frac{(1-t_1+k-s)R}{1-t_2} - (1+r)(1+k)$  with  $\tau = k = 0$ ) as the education subsidy takes over the role of the tax on unskilled labor in offsetting the learning distortion imposed by the tax on skilled labor. If education expenses  $k$  are not tax deductible, the optimal labor income tax may be regressive. Indeed, in the absence of a capital income tax, the tax rate on skilled labor is given by  $t_1 = (1+k) t_2 - s$  if first-period leisure demand is exogenous (see (22) and use the definition of  $\Delta \equiv \frac{(1-t_1+k-s)R}{1-t_2} - (1+r)(1+k)$  with  $\tau = s = 0$ ). Accordingly, the optimal labor tax is regressive if education subsidies are small, non-deductible expenses are important, and distributional considerations are important (so that the optimal tax on skilled labor is large). Intuitively, if the labor tax does not allow deductibility of education expenses, the tax on unskilled labor becomes a less effective instrument to boost learning. Hence, the tax on unskilled labor needs to be raised more to offset the learning distortions on account of the redistributive tax on skilled labor.

Even in the absence of capital income taxes and education subsidies, the labor income tax is progressive if first-period leisure demand is endogenous

( $E_{1t_1}, E_{1t_1}^* > 0$ ). Elastic first-period leisure demand models the concerns of many policymakers that taxes on unskilled labor harm the incentives of unskilled workers to seek employment. These concerns strengthen the case for progressive labor taxes. The reason is that with endogenous first-period leisure demand a tax on unskilled labor induces agents to spend more time not only learning but also enjoying leisure. In this way, the tax not only corrects for inadequate incentives to accumulate human capital, but also induces excessive leisure demand so that the tax implies both favorable and unfavorable substitution effects. As a direct consequence, the government no longer has access to a non-distortionary instrument to offset the learning distortion implied by the tax on skilled labor. It thus has to trade off distortions in learning against distortions in first-period leisure demand. This implies that human capital accumulation is distorted in the optimum (see (26) with  $E_{1t_1}^* > 0$ ). Moreover, *ceteris paribus* the distributional characteristic  $\xi$ , the government optimally sets a smaller tax rate on skilled labor  $t_2$  than with exogenous leisure demand because it no longer can costlessly offset the learning distortions implied by this tax (see (24) and note that  $\varepsilon^*$  rises with  $E_{1t_1}^*$ ). Indeed, (24) implies that the optimal tax on skilled labor declines with the learning elasticity  $\beta$  and the elasticities of leisure demands in both periods (i.e.  $\varepsilon_1$  and  $\varepsilon_2$ ).

**Proposition 1** (Optimal labor income taxation) *The optimal labor tax is progressive ( $t_2 > t_1$ ) if first-period leisure demand is elastic ( $\varepsilon_1 > 0$ ), the capital income tax is positive ( $\tau > 0$ ), or education subsidies are positive ( $s > 0$ ). The tax system is flat ( $t_1 = t_2$ ) if first-period leisure demand is inelastic ( $\varepsilon_1 = 0$ ), the capital income tax is zero ( $\tau = 0$ ), and if direct costs are deductible ( $s = t_1 k$ ). The optimal labor tax structure eliminates tax distortions in learning ( $\Delta = 0$ ) only if first-period leisure demand is exogenous ( $\varepsilon_1 = 0$ ) or redistributive motives are absent ( $\xi = 0$ ).*

## 6 Optimal dual income taxation

Until now we have assumed that the capital income tax  $\tau$  was exogenously fixed. This section allows the government to freely employ this tax to optimize social welfare. This allows us to investigate the optimal mix between capital income taxation and a labor income schedule with two brackets.

## 6.1 Optimal capital income tax

The first-order condition for the optimal tax on capital income  $\tau$  is given by (using Roy's identity  $\partial v_\alpha / \partial \tau = r a_\alpha = r[(1-t_1)(1-h_1) - c_1 - (1-t_1+k-s)e_\alpha]$ ):

$$\begin{aligned} & \int_{\underline{\alpha}}^{\infty} (b_\alpha - 1) a_\alpha dF - \Delta \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{r \partial \tau} dF \\ &= -\tau r \int_{\underline{\alpha}}^{\infty} \frac{\partial c_1}{r \partial \tau} dF - (t_1 R + \tau r) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1}{r \partial \tau} dF + \int_{\underline{\alpha}}^{\infty} t_2 \phi(\alpha; e_\alpha) \frac{\partial l_2}{r \partial \tau} dF. \end{aligned}$$

Substituting (1) to eliminate  $a_\alpha$  and (11) to eliminate  $l_{2\alpha} \phi(\alpha; e_\alpha)$  while using (17), we arrive at:

$$\begin{aligned} & \xi_e (1 - t_1 + k - s) \int_{\underline{\alpha}}^{\infty} e_\alpha dF + \frac{\tau r}{R} \varepsilon_{c_1 \tau} c_1 \\ &= - \left( t_1 + \frac{\tau r}{R} \right) \varepsilon_{(1-h_1)\tau} (1 - h_1) + \frac{\Delta}{R} \varepsilon_{e\tau} \int_{\underline{\alpha}}^{\infty} e_\alpha dF + \frac{t_2}{R} \mu \varepsilon_{l_2 \tau} \int_{\underline{\alpha}}^{\infty} e_\alpha dF, \end{aligned} \quad (27)$$

since the following elasticities are independent of skill:  $\varepsilon_{e\tau} \equiv \frac{\partial e}{r \partial \tau} \frac{R}{e} = \frac{1}{1-\beta(1+\varepsilon_2)}$  (the second equality follows from (9)),  $\varepsilon_{c_1 \tau} \equiv \frac{\partial c_1}{r \partial \tau} \frac{R}{c_1} = \varepsilon_c$  (the second equality follows from (6)),  $\varepsilon_{(1-h_1)\tau} \equiv \frac{\partial h_1}{r \partial \tau} \frac{R}{(1-h_1)} = \varepsilon_1$ , and  $\varepsilon_{l_2 \tau} \equiv \frac{\partial l_2}{r \partial \tau} \frac{R}{l_2} = \frac{\beta \varepsilon_2}{1-\beta(1+\varepsilon_2)}$ . Expression (27) shows that, at the optimum capital income tax, the marginal costs of raising the tax should equal the benefits of doing so. The three terms at the left-hand side of the expression stand for the costs: distributional losses (as the lifetime poor save more because they learn less and thus concentrate more of their work effort in the beginning of their lives) and the worsening of distortions in the intertemporal allocation of consumption (assuming that  $\tau > 0$ ) and first-period leisure demand (assuming that  $(t_1 + \tau r/R) > 0$ ). The two terms on the right-hand side of (27) represent the benefits of raising the capital income tax, namely first-order welfare gains due to more learning (assuming that  $\Delta > 0$ ) and higher skilled labor supply (assuming that  $t_2 > 0$ ).

In the absence of other taxes and subsidies (i.e.  $t_1 = t_2 = s = 0$  so that  $\Delta = -(1+k)\tau r$ ), capital income is subsidized:

$$\frac{\tau r}{R} = \frac{-\xi}{\left[ \varepsilon_c \frac{c_1}{(1+k) \int_{\underline{\alpha}}^{\infty} e_\alpha dF} + \varepsilon_1 \frac{1-h_1}{(1+k) \int_{\underline{\alpha}}^{\infty} e_\alpha dF} + \frac{1}{1-\beta(1+\varepsilon_2)} \right]}. \quad (28)$$

This subsidy helps to alleviate the inequities in lifetime incomes. Indeed, the subsidy rises with the distributional characteristic  $\xi$ . The three terms in the denominator of (28) correspond to the three decision margins that are distorted by the capital income tax: first-period leisure and consumption demands and learning.

## 6.2 Optimal taxes on capital and unskilled labor

If the government can freely set  $\tau$  and  $t_1$ , we can substitute (18) into (27) to eliminate the distributional characteristic  $\xi$  to arrive at (by using  $E_{1t_1} = \frac{\varepsilon_{(1-h_1)\tau(1-h_1)}}{\int_{\alpha}^{\infty} e_{\alpha} dF}$ ,  $\frac{\varepsilon_{e\tau}}{(1-t_1+k-s)} = \frac{\varepsilon_{et_1}}{(1-t_1)}$ , and  $\frac{\varepsilon_{l_2\tau}}{(1-t_1+k-s)} = \frac{\varepsilon_{l_2t_1}}{(1-t_1)}$ ):

$$(k-s) \left( \frac{t_1 + \frac{\tau r}{R}}{1-t_1} \right) \varepsilon_1 (1-h_1) = \frac{\tau r}{R} \varepsilon_{c_1\tau} c_1. \quad (29)$$

If non-deductible expenses and education subsidies are absent (i.e.  $k = s = 0$ ), first-period leisure demand is inelastic (i.e.  $\varepsilon_1 = 0$ ) and taxes on skilled labor are not available to pursue distributional objectives (i.e.  $t_2 = 0$ ), unskilled labor is subsidized (see (19)) while capital remains untaxed. If not all education expenses are deductible to determine the base of the unskilled labor subsidy (i.e.  $k > 0$  while  $s = 0$ ), the capital subsidy becomes a more efficient instrument to redistribute in favor of the poor as the higher educational expenses of the able no longer narrow the base of the labor subsidy. This distributional benefit of the capital income subsidy has to be weighted against the additional distortion of the capital income subsidy on the intertemporal allocation of consumption.

## 6.3 Optimal dual income tax

If the government can freely determine all parameters of a dual income tax (i.e.  $g$ ,  $t_1$ ,  $t_2$ , and  $\tau$ ), we can combine (22) and (29). The optimal tax on capital income is zero if first-period leisure demand is exogenous (see (29) with  $\varepsilon_1 = 0$ ) or all education expenses are deductible from the labor-income tax in the first period.<sup>28</sup> With exogenous first-period leisure demand, the optimal labor tax is flat (i.e.  $t_1 = t_2$ ). Hence, the learning decision is not

<sup>28</sup>With deductible expenses, (24), (25), and (26) determine the optimal tax structure with  $\tau = 0$  (see the appendix).



distorted, so that there is no role for capital income taxation in offsetting this distortion. With endogenous first-period leisure demand, the optimal labor tax is progressive, thereby discouraging agents to accumulate human capital. If all education expenses are deductible, however, the capital income tax still plays no role in offsetting the learning distortion. The reason is that the first-period tax rate on labor is a more efficient instrument to stimulate learning. Compared to the capital income tax, the tax on unskilled labor imposes the same distortions on first-period leisure demand but, in contrast to the capital income tax, it does not distort the intertemporal allocation of consumption.<sup>29</sup>

The optimal capital income tax is positive if not all education expenses are deductible from the labor income tax (i.e.  $k > 0$ ,  $s = 0$ ) and first-period leisure demand is endogenous (see (29) with  $k - s > 0$  and  $\varepsilon_1 > 0$ ). Intuitively, if the labor tax does not allow education expenses to be deducted, the first-period labor tax becomes a less effective instrument to stimulate learning. As a direct consequence, the tax on unskilled labor would have to be raised substantially to offset the learning distortion due to the tax on skilled labor  $t_2 > 0$ . Such a large first-period labor tax would impose serious distortions in the first-period labor market by encouraging agents to substitute leisure for working time. To contain these distortions in the labor market, the government relies on the capital income tax to offset the learning distortions on account of the redistributive tax on skilled labor. In contrast to the tax on unskilled labor, however, the capital income tax distorts the intertemporal allocation of consumption. Hence, the optimal mix of taxes on unskilled labor and capital income balance distortions in the labor and capital markets.

We thus find conditions for a dual income tax with a positive capital income tax and a progressive labor tax to be optimal as a redistributive tax system.<sup>30</sup> The tax on capital income plays an important role in the optimal

---

<sup>29</sup>The result that the optimal capital income tax is zero if all education expenses are tax deductible depends on the specific utility function (4) in which leisure demands are weakly separable from consumption. If leisure demands would not be separable from the intertemporal allocation of consumption, the government would like to employ the capital income tax to reduce distortions in the labor market.

<sup>30</sup>The optimal capital income tax eliminates learning distortions if first-period consumption demand is inelastic (i.e.  $\varepsilon_c = 0$ ). A subsidy on unskilled labor then offsets the distortions on first-period labor supply:  $t_1 = -\tau r/R$ . If  $\varepsilon_c, \varepsilon_1 > 0$ , the learning distortion is positive in equilibrium (if  $s = 0$ ).

tax system and the labor tax becomes more progressive if the capital income tax is a relatively efficient instrument to boost learning (because it does not distort the intertemporal allocation of consumption much, i.e.  $\varepsilon_c$  is small) while the tax on unskilled labor is relatively inefficient (because it encourages the unskilled to substantially raise leisure demand, i.e.  $\varepsilon_1$  is large, and it does not allow substantial education expenses to be deducted). The tax on capital also becomes a more important tax instrument if distributional considerations become more prominent. This may seem counterintuitive because the capital income tax is actually a regressive tax since the low skilled save more than the high skilled do. The reason why this regressive tax is nevertheless used more intensively if the government wants to redistribute more is that the tax on skilled labor is a more efficient instrument for redistribution than the capital income tax, which is targeted solely at offsetting the learning distortions from the redistributive tax on skilled labor. Indeed, a positive capital income tax allows for a more progressive labor tax system. Education subsidies reduce the potential role of capital taxes in offsetting learning decisions, but also make the labor tax system more progressive.

**Proposition 2** (Optimal dual income taxation) *The optimal tax on capital income is zero ( $\tau = 0$ ) if first-period leisure demand is exogenous ( $\varepsilon_1 = 0$ ) or all education expenses are deductible from the labor-income tax in the first period ( $s = t_1 k$ ). If not all education expenses are deductible from the labor income tax ( $s < t_1 k$ ) and first-period leisure demand is endogenous ( $\varepsilon_1 > 0$ ), the optimal tax on capital income is positive ( $\tau > 0$ ). The capital income tax becomes larger and the labor tax schedule becomes more progressive if the intertemporal substitution effects in consumption are small ( $\varepsilon_c$  small), unskilled labor supply is elastic ( $\varepsilon_1$  large), and education subsidies are low ( $s$  small).*

## 7 Education subsidies

This section allows the government to employ the instrument of education subsidies to optimize social welfare. This allows us to investigate the impact of this additional policy instrument on the optimal tax structure.

## 7.1 Optimal education subsidies

The first-order condition for education subsidies  $s$  amounts to (using  $\frac{\partial c_1}{\partial s} = \frac{\partial h_1}{\partial s} = 0$  (see (6) and (7), respectively) and Roy's identity  $\partial v_\alpha / \partial s = R e_\alpha$ ):

$$\int_{\underline{\alpha}}^{\infty} (b_\alpha - 1) e_\alpha dF + \frac{\Delta}{R} \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{\partial s} dF + \int_{\underline{\alpha}}^{\infty} \left( \frac{t_2 \phi(\alpha; e_\alpha)}{R} \frac{\partial l_2}{\partial s} \right) dF = 0.$$

Employing (17) and using (11) to eliminate economy-wide labor income in the second period, we find:

$$\xi_e = \frac{\Delta}{R} \frac{\varepsilon_{es}}{(1-s)} + \frac{t_2}{R} \frac{\mu \varepsilon_{l_2 s}}{(1-s)}, \quad (30)$$

where we have used  $\varepsilon_{es} \equiv \frac{\partial e}{\partial s} \frac{(1-s)}{e} = \frac{(1-s)}{(1-\beta(1+\varepsilon_2))(1-t_1+k-s)}$  (the last equality follows from (9)) and  $\varepsilon_{l_2 s} \equiv \frac{\partial l_2}{\partial s} \frac{(1-s)}{l_2} = \frac{\beta \varepsilon_2 (1-s)}{(1-\beta(1+\varepsilon_2))(1-t_1+k-s)}$  (the last equality follows from (8)), which do not depend on skill. The distributional costs of higher educational benefits (i.e. the left-hand side of (30)) should correspond to the additional first-order welfare benefits of the higher subsidies (i.e. the right-hand side of (30)). These welfare benefits consist of the impact on the learning distortion  $\Delta$  (assuming  $\Delta > 0$ ) and the distortion in second-period labor supply  $t_2$  (assuming  $t_2 > 0$ ). In case taxes are absent (i.e.  $t_1 = t_2 = \tau = 0$  so that  $\Delta/R = -s$ ), we can employ (30) to solve for  $s$  (by using the definitions of the elasticities  $\varepsilon_{es}$ ):

$$\frac{s}{1+k-s} = -\xi(1-\beta(1+\varepsilon_2)). \quad (31)$$

Without any other instruments to tax skill, taxes on education (i.e.  $s < 0$ ) are used to reduce inequities in lifetime incomes.

## 7.2 Optimal tax on unskilled labor

If the government can freely set not only the education subsidies but also the tax on unskilled labor  $t_1$ , we can substitute (30) into (18) to eliminate the distributional characteristic  $\xi_e$  to find that (using  $\frac{\varepsilon_{es}}{(1-s)} = \frac{\varepsilon_{et_1}}{(1-t_1)}$  and  $\frac{\varepsilon_{l_2 s}}{(1-s)} = \frac{\varepsilon_{l_2 t_1}}{(1-t_1)}$ ) the distortion in first-period leisure demand is zero (i.e.  $t_1 + \tau r / R = 0$ ). An optimal subsidy on unskilled labor:

$$t_1 = -\frac{\tau r}{R},$$

thus corrects for the impact of a positive capital income tax on the consumption of first-period leisure. Without an education subsidy, the tax on unskilled labor could not be aimed solely at offsetting the distortions on first-period labor supply since this tax must be used also to offset the learning distortions implied by the redistributive tax on skill. An optimal education subsidy takes care of the second task so that the first-period tax can be targeted at removing distortions in the labor market in the first period.

### 7.3 Optimal taxes on unskilled labor and capital

If the government can also set the capital income tax optimally, expression (29) implies that both the capital income tax and the tax on unskilled labor are absent, i.e.:

$$\tau = t_1 = 0.$$

Intuitively, a tax on education is a more efficient instrument to tax skill than subsidies on capital income or unskilled labor are because a tax on education does not distort first-period leisure demand. In this case, the optimal education subsidy is given from (30) (with  $\tau = t_1 = 0$ ):

$$\frac{s}{1+k-s} = \frac{t_2}{1-t_2}(1+\varepsilon_2) - \xi(1-\beta(1+\varepsilon_2)).$$

Despite their adverse distributional consequences, education subsidies may be used to offset the adverse impact of second-period taxes (i.e.  $t_2 > 0$ ) on learning and on skilled labor supply. Indeed, *ceteris paribus* the distributional characteristic  $\xi$ , education subsidies rise with the elasticities of learning  $\beta$  and second-period leisure demand  $\varepsilon_2$  (assuming that  $t_2 > 0$ ).

### 7.4 Optimal tax on skilled labor

If the government can simultaneously tax skilled labor supply and provide education subsidies, we find (from substituting (30) into (20) to eliminate the distribution characteristic  $\xi_e = \xi_2$  and using that  $\frac{\mu\varepsilon_{l_2s}}{R(1-s)} = \frac{\varepsilon_{l_2t_2}}{(1-t_2)}$  and  $\frac{\varepsilon_{es}}{R(1-s)} > \frac{\varepsilon_{et_2}}{\mu(1-t_2)})$  that education subsidies are set so that the learning wedge  $\Delta$  is zero. Taxes on skill are aimed at reducing inequities, while the education subsidies eliminate distortions in learning. Irrespective of the income taxes in the first period  $t_1$  and  $\tau$ , the optimal tax on skill is given by:

$$\frac{t_2}{1-t_2} = \frac{\xi}{\varepsilon_{l_2t_2}} = \frac{\xi(1-\beta(1+\varepsilon_2))}{\varepsilon_2}. \quad (32)$$

Comparing (32) with (21), we observe that the additional instrument of the education subsidy (*ceteris paribus* the distributional characteristic  $\xi$ ) allows for a higher tax on skilled labor. Intuitively, education subsidies offset the learning distortions implied by the tax on skill, so that the tax on skilled labor distorts only second-period labor supply. In this way, the combination of the tax on skilled labor and the education subsidy allows the government to tax the inframarginal rents (i.e. pure profits) from learning without distorting the marginal incentives to learn.<sup>31</sup> This improves the trade-off between equity and efficiency considerations. Indeed, in equilibrium, the presence of optimal education subsidies allows the government to be more successful in combatting inequities, which will reduce the distributional characteristic  $\xi$ .

In order to ensure that learning is not distorted (i.e.  $\Delta = 0$ ), the optimal education subsidy amounts to:

$$s = (1 + k) \left( 1 - \frac{(1 - t_2)(1 + r)}{(1 + r(1 - \tau))} \right) - t_1. \quad (33)$$

where  $t_2$  is given from (32), while the other taxes  $\tau$  and  $t_1$  are exogenously given.<sup>32</sup> In order to interpret this expression, we consider several special cases in turn.

The first case assumes that the capital income tax is zero ( $\tau = 0$ ), while direct costs of education are absent ( $k = 0$ ). In that case, the optimal education subsidy corrects for learning distortions on account of the progression in the labor income tax (i.e.  $s = t_2 - t_1$ ). If the tax system is flat (i.e.  $t \equiv t_2 = t_1$ ) and the capital income tax is zero ( $\tau = 0$ ), the optimal subsidy offsets the impact of non-deductible education expenses on the incentive to learn (i.e.  $s = tk$ ). This implies that these expenses have effectively become deductible against the labor income tax. A third special case is when education expenses are absent ( $k = 0$ ) while the tax system is flat (i.e.  $t \equiv t_2 = t_1$ ). In that case, a positive capital income tax implies that, in addition to financial saving, also human capital accumulation must be taxed on a net basis (i.e.  $s = -\frac{r\tau(1-t)}{1+r(1-\tau)} < 0$ ) in order to prevent a distortion in the portfolio choice between human and financial capital.

---

<sup>31</sup>Note that endogenous learning (i.e.  $\beta > 0$ ) still reduces the optimal tax on skilled labor by raising the elasticity of second-period labor supply  $\varepsilon_2/(1 - \beta(1 + \varepsilon_2))$ .

<sup>32</sup>This expression continues to hold if  $\tau$  and  $t_1$  are set optimally in the presence of general preferences (see the appendix).

## 7.5 Optimal dual income tax

If the government can optimally set all tax instruments and education subsidies, the tax in skill  $t_2$  is given (32), while unskilled labor and capital income remain untaxed. Hence, a special dual tax system emerges: labor income is taxed at a rate  $t_2$  with a large tax-free allowance  $\chi$ , whereas capital income is tax free. The labor tax thus resembles an earned income tax credit (EITC), where labor income of unskilled labor goes largely untaxed.

The instrument of education subsidies thus has important effects on the optimal dual income tax. In particular, it makes the labor tax more progressive by raising  $t_2$  and reducing  $t_1$ . The elasticity of first-period leisure demand determines the contributions of a higher  $t_2$  and a lower  $t_1$  to the more progressive labor tax. Elastic first-period labor demand tends to raise the positive response of the tax on skilled labor to the availability of education subsidies (compare (32) with (24)). Intuitively, education subsidies are more efficient than taxes on unskilled labor in ensuring that inframarginal rents on learning are taxed because they do not distort the demand of leisure by unskilled workers. The tax on unskilled labor is thus reduced, thereby making the tax system more progressive. A more sensitive leisure margin in the first period reduces the drop in the unskilled tax rate in response to the availability of education subsidies because it makes the unskilled tax rate an unattractive instrument to correct the learning margin even if education subsidies are absent.

With separable preferences, the presence of education subsidies eliminates the case for a positive capital income tax as an instrument to stimulate learning in the presence of non-deductible education expenses and endogenous first-period leisure. The reason is that education subsidies are a more efficient instrument to deal with the learning distortion than capital income taxes, which distort also first-period consumption of leisure and commodities. In the presence of additively separable preferences, we thus no longer have a case for a positive capital income tax. The education subsidies ensure that neither human capital investment nor financial investment are distorted, even though the labor tax is progressive.<sup>33</sup>

In the absence of first-period income and capital taxes, the optimal edu-

---

<sup>33</sup>The optimal zero capital income, however, tax depends crucially on the assumed additive preference structure and the availability of government debt as an instrument to correct the intergenerational distribution of resources (see Bernheim (1999)).

cation subsidy is given by:

$$s = (1 + k)t_2 = \frac{(1 + k)\xi(1 - \beta(1 + \varepsilon_2))}{\varepsilon_2 + \xi(1 - \beta(1 + \varepsilon_2))}. \quad (34)$$

This expression reveals that education subsidies become more important if distributional concerns become more prominent, as indicated by a larger distributional characteristic  $\xi$ . Clearly, education subsidies and redistribution are Siamese twins since the government employs education subsidies to offset the adverse impact of taxes on the incentives to accumulate human capital. Indeed, education subsidies would be zero if redistributional considerations would be absent (i.e.  $\xi = 0$ ).

**Proposition 3** (Optimal education subsidies) *Optimal education subsidies ensure that investment in human capital is efficient in a world where distortionary taxes are used to generate revenues and to redistribute incomes. The optimal subsidy increases with distributional concerns ( $\xi$  large) and non-deductible direct costs of education ( $k$  large). It decreases with the elasticity of skilled labor supply  $\varepsilon_2$  and learning  $\beta$ . If preferences are separable, optimal education subsidies allow capital income and unskilled labor income to go untaxed (i.e.  $\tau = t_1 = 0$ ) while skilled labor income is taxed ( $t_2 > 0$ ).*

## 8 Are education subsidies optimally set?

This section explores whether the current levels of education subsidies in several OECD countries are efficient. To compute optimal education subsidies, we employ (33) and use observed values of the tax parameters  $t_1$ ,  $t_2$ , and  $\tau$ . We then compare these optimal subsidies with the actually observed education subsidies in several OECD countries. We confine ourselves to subsidies to higher education because compulsory schooling laws ensure that progressive taxes do not reduce participation in basic education.

Table A1 in the Appendix contains the required data.<sup>34</sup> Since the data sources do not fully cover all OECD countries, we limit our calculations

---

<sup>34</sup>Our measures for wages and subsidies apply to 1997 and 1995 while the tax figures apply to 1997 and 2000. Since education policies and tax schedules are rather stable over time, this should not cause serious problems.

to eight countries: Canada, Denmark, Finland, Germany, Italy, the Netherlands and the United States. We employ the marginal rate in the tax bracket containing foregone earnings (minus general exemptions) and including local taxes as a measure for  $t_1$ . Similarly, the marginal tax rate in the tax bracket containing the income of a college educated worker (minus exemptions) and including local taxes is used as a measure for  $t_2$ . As a proxy for foregone earnings, we take the average yearly gross wage of a male worker with less than 15 years of education. The wage of an educated worker is the average yearly gross wage of a male worker with more than 15 years of education. With this definition, 20% of the overall sample consists of higher educated workers.<sup>35</sup> Internationally comparable data on earnings are taken from the International Adult Literacy Survey by OECD/Statistics Canada (1995). Data from the International Bureau for Fiscal Documentation (IFBD, 1997; IFBD, 2000) provide the required information on statutory income tax structures.

As regards the effective tax rate on capital income, countries tax various sources of capital income in a non-uniform fashion. In order to capture the potential influence of capital income taxes on learning decisions, we employ the average effective rate on capital income as reported in OECD (2000b). In addition, we present calculations in which capital taxes are set at zero. The term  $\frac{(1+r)}{(1+r(1-\tau))}$  in (33) from our two-period model is derived from the effective yearly interest rate  $\delta$  by assuming that students are enrolled for five years in education between ages  $a^o = 18$  and  $a' = 23$  and then enter the labor market until the age of  $a^* = 65$ . Hence, the first-order condition for optimal learning reads as follows:

$$\begin{aligned} (1 - t_2)l_2\phi_e(.) \int_{a'}^{a^*} \exp[-\delta(1 - \tau)v]dv \\ = (1 - t_1 + k - s) \int_{a^o}^{a'} \exp[-\delta(1 - \tau)v]dv, \end{aligned}$$

where the left-hand side corresponds to the marginal return from the human capital investment (assumed to be constant over time). The right-hand side measures the discounted value of the learning costs. Straightforward

---

<sup>35</sup>We thus assume that income differences between skilled and unskilled workers are attributable to education. Table A1 in the Appendix reveals that the returns on education implied by our assumption are close to micro-econometric estimates correcting for ability bias (see e.g. Ashenfelter et al. (2000)).



manipulation yields the following:

$$\frac{1+r}{1+r(1-\tau)} = \left( \frac{\exp[\delta(a' - a^o)] - 1}{1 - \exp[\delta(a' - a^*)]} \right) \left( \frac{1 - \exp[\delta(1-\tau)(a' - a^*)]}{\exp[\delta(1-\tau)(a' - a^o)] - 1} \right).$$

Our calculations assume an effective interest rate of 2% per annum (i.e.  $\delta = .02$ ).

If the first-period wage rate (measuring forgone earnings) are  $w_1$  rather than unity (as assumed in the theoretical model), the optimal subsidy formula reads as:<sup>36</sup>

$$\frac{s^*}{w_1} = \left( 1 + \frac{k}{w_1} \right) \left( 1 - (1 - t_2) \frac{(1+r)}{1+r(1-\tau)} \right) - t_1.$$

We measure the subsidy  $\frac{s^*}{w_1}$  as total public expenditure per student per year divided by gross foregone earnings. Direct costs  $\frac{k}{w_1}$  are total (i.e. public and private) direct expenditures on education divided by gross foregone earnings. Actual subsidies and expenditures on higher education are provided by OECD (2000a). The wage data are denominated in the various currencies. Therefore, we used the OECD PPP-deflator to transform the education expenditures and subsidies back to original currencies.

Table 1 contains the actual and optimal education subsidies as a percentage of foregone earnings. Actual subsidies are relatively low in Denmark because forgone earnings are high due to a compressed wage distribution. The reverse holds true for the United States and Canada, where low-skilled wages are relatively low. Even if we employ the positive average effective rates on capital income as reported in OECD (2000b), optimal education subsidies are positive and range from 6-27%. Hence, a significant part of actual education subsidies can be justified on pure efficiency grounds. Without any positive capital income taxes<sup>37</sup>, optimal education subsidies are considerably larger. In fact, for most countries, the optimal subsidies are close to the actually observed subsidies. In view of the problematic measurement of

---

<sup>36</sup>This expression, which contains only policy variables without any preference parameters, requires only that  $s$  and  $t_2$  are set optimally. Indeed, the appendix shows that this expression holds for a general utility function with arbitrary (i.e. not necessarily optimal) values for the other policy parameters  $t_1$  and  $\tau$ .

<sup>37</sup>Gordon and Slemrod (1988) argue that this is in fact the relevant case. In their view, observed tax revenues from capital income taxes are in fact taxes on rents. With preferences that are intertemporally separable, zero capital income taxes would actually be optimal in our model.

capital income tax, the two alternative calculations of optimal subsidies serve as lower and upper bounds for the optimal subsidies.

Table 1 - Actual and optimal subsidies (% of forgone earnings)

	actual	optimal	optimal ( $\tau = 0$ )
Canada	50	25	42
Denmark	25	25	32
Finland	37	10	16
Germany	39	13	23
Italy	29	6	18
Netherlands	40	27	35
Sweden	75	7	24
United States	43	23	40

Sources: see appendix.

## 9 Conclusion

This paper has studied the optimal setting of a progressive labor tax, a flat capital income tax and education subsidies. We showed that education subsidies are a powerful instrument to eliminate distortions in the accumulation of human capital associated with redistributive policies favoring the unskilled. Education subsidies and redistribution of incomes are thus like Siamese twins – even though the ones who benefit from the subsidies enjoy relatively high lifetime incomes. The more the government desires to help the unable, the more it should employ education subsidies to offset the learning distortions associated with redistribution. We showed that a substantial part of existing education subsidies in some important OECD countries can be justified on these grounds.

In the absence of education subsidies, we demonstrated that the capital income tax component of a dual income tax may be positive, even if preferences are separable. With these preferences, however, a positive capital income tax requires the presence of non-deductible education expenses. We showed also that the labor income tax component of a dual income tax is progressive, even if skilled labor supply is relatively elastic compared to unskilled labor supply. The introduction of education subsidies reduces the case for a positive capital income tax, but increases the optimal progression of the labor income tax.

In future research we would like to introduce non-tax distortions, such as positive externalities from human capital accumulation, liquidity constraints, wage rigidities, and other labor-market distortions (such as union power and efficiency wages). Since these non-tax distortions can be viewed as implicit tax wedges, the results in this paper provide already some insights. In particular, positive externalities from human capital accumulation can be viewed as an implicit tax on learning, liquidity constraints as an implicit subsidy on capital income, and downward wage rigidities and union power as an implicit tax on labor. Furthermore, following van Ewijk and Tang (2000), we may introduce non-verifiable training efforts so that education subsidies are no longer a costless instrument to alleviate learning distortions. The results in this paper in which education subsidies are exogenously given rather than optimally set provide already some insights in this case. In particular, if training efforts are completely non-verifiable, the cases without any education subsidies become relevant.

## References

- Acemoglu, Daron J., and Joshua Angrist (1999), “How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws”, NBER Working Paper 7444.
- Anersson, Fredrik, and Kai A. Konrad (2000), “Human Capital Investment and Globalization in Extortionary States”, mimeo.
- Ashenfelter, Orley, Colm Harmon and Hessel Oosterbeek (1999), “A Review of Estimates of the Schooling/Earnings Relationship with Tests for Publication Bias”, *Labour Economics* 6, 453-470.
- Atkinson, Anthony B., and Agnar Sandmo (1980), “Welfare Implications of the Taxation of Savings”, *Economic Journal* 90, 529-549.
- Atkinson, Anthony B., and Joseph E. Stiglitz (1980), *Lectures on Public Economics*, New York: McGraw-Hill.
- Bernheim, B. Douglas (1999), “Taxation and Saving”, NBER Working Paper 7061.
- Boskin, Michael (1977), “Notes on the Tax Treatment of Human Capital”, NBER Working Paper 116.

- Boadway, Robin, Nicolas Marceau, and Maurice Marchand (1996), "Investment in Education and the Time Inconsistency of Redistributive Tax Policy", *Economica* 63, 171-189.
- Bovenberg, A. Lans, and Casper van Ewijk (1997), "Progressive Taxes, Equity, and Human Capital Accumulation in an Endogenous Growth Model with Overlapping Generations", *Journal of Public Economics* 64, 154-179.
- Cameron, Stephen, and Christopher Taber (2000), "Borrowing Constraints and the Returns to Schooling", NBER Working Paper 7761.
- Chamley, Christophe (1986), "Optimal Taxation of Capital Income in General Equilibrium With Infinite Lives", *Econometrica* 54 (3), 607-622.
- Diamond, Peter A. (1998), "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review* 88 (1), 83-95.
- Diamond, Peter A., and James A. Mirrlees (1971), "Optimal Taxation and Public Production I: Production Efficiency", *American Economic Review* 61, 8-27.
- Driffil, E. John, and Harvey S. Rosen (1983), "Taxation and Excess Burden: A Life Cycle Perspective", *International Economic Review* 3, 671-683.
- Dur, Robert, and Coen N. Teulings (2001), "Education and Efficient Redistribution", mimeo: Erasmus University Rotterdam.
- Eaton, Jonathan, and Harvey S. Rosen (1980), "Taxation, Human Capital and Uncertainty", *American Economic Review* 70 (4), 705-715.
- Ewijk, Casper, and Paul J.G. Tang (2000), "Efficient Progressive Taxes and Education Subsidies", CPB Research Memorandum 170, CPB Netherlands Bureau for Economic Policy Analysis.
- Gordon, Roger H., and Joel Slemrod (1988), "Do We Collect any Revenue from Taxing Capital Income," *Tax Policy and the Economy* 2, 89-130.
- Hare, Paul G., and David T. Ulph (1979), "On Education and Distribution", *Journal of Political Economy* 87, S193-S212.

- Heckman, James J. (1976), "A Life-Cycle Model of Earnings, Learning and Consumption", *Journal of Political Economy* 4, S11-S44.
- Heckman, James J., and Peter J. Klenow (1997), "Human Capital Policy", mimeo: University of Chicago.
- Hendricks, Lutz (1999), "Taxation and Long-Run Growth", *Journal of Monetary Economics* 43, 411-434.
- International Bureau of Fiscal Documentation (1997), *European Tax Handbook 1997*, Amsterdam: IBFD Publications BV.
- International Bureau of Fiscal Documentation (2000), *European Tax Handbook 2000*, Amsterdam: IBFD Publications BV.
- Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi (1993), "Optimal Taxation in Models of Endogenous Technological Change", *Journal of Political Economy* 101 (3), 485-517.
- Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi (1997), "On the Optimal Taxation of Capital Income", *Journal of Economic Theory* 73, 93-117.
- Judd, Kenneth L. (1999), "Optimal Taxation and Spending in General Competitive Growth Models", *Journal of Public Economics* 71, 1-26.
- King, Mervyn A. (1980), "Savings and Taxation", in: G. A. Hughes and G. M. Heal (eds.), *Public Policy and the Tax System*, London.
- King, Mervyn A., and Don Fullerton (1984), *The Taxation of Income from Capital: A Comparative Study of the United States, The United Kingdom, Sweden, and West-Germany*, Chicago: Chicago University Press.
- Kotlikoff, Laurence J., and Lawrence H. Summers (1979), "Tax Incidence in a Life Cycle Model with Variable Labor Supply", *Quarterly Journal of Economics* 93, 705-718.
- Krueger, Alan B., and Mikael Lindahl (1999), "Education and Growth: Why and for Whom", NBER Working Paper 7591.
- Lommerud, Kjell Erik (1989), "Educational Subsidies when Relative Income Matters", *Oxford Economic Papers* 41, 640-652.

- Lord, William (1989), “The Transition from Payroll to Consumption Receipts with Endogenous Human Capital“, *Journal of Public Economics* 38, 53-73.
- Milesi-Ferretti, Gian Maria, and Nouriel Roubini (1998), “On the Taxation of Human and Physical Capital in Models of Endogenous Growth”, *Journal of Public Economics* 70, 237-254.
- Mirrlees, James A. (1971), “An Exploration in the Theory of Optimum Income Taxation”, *Review of Economic Studies* 38, 175-208.
- Nerlove, Marc, Assaf Razin, Efraim Sadka, Robert K. von Weizsäcker (1993), “Comprehensive Income Taxation, Investments in Human and Physical Capital, and Productivity”, *Journal of Public Economics* 50, 397-406.
- Nielsen, Søren B., and Peter B. Sørensen (1997), “On the Optimality of the Nordic System of Dual Income Taxation”, *Journal of Public Economics* 63, 311-329.
- OECD (1996), *Education at a Glance*, Paris: OECD.
- OECD (2000a), *Education at a Glance*, Paris: OECD.
- OECD (2000b), “Average Effective Tax Rates on Capital, Labour and Consumption”, by David Carey and Harry Tchilinguirian, OECD Economics Department Working Paper 258, Paris: OECD.
- OECD/Statistics Canada (1995), *Literacy, Economy and Society: Results of the First International Adult Literacy Survey*, Paris: OECD.
- Ordoover, J.A., and Edmund S. Phelps (1979), “The Concept of Optimal Taxation in the Overlapping-Generations Model of Capital and Wealth”, *Journal of Public Economics* 12, 1-26.
- Pecorino, Paul (1993), “Tax Structure and Growth in a Model with Human Capital”, *Journal of Public Economics* 52, 251-271.
- Sandmo, Agnar (1985), “The Effects of Taxation on Saving Risk Taking”, in: Alan Auerbach and Martin Feldstein (eds.) (1985), *Handbook of Public Economics - Vol. I*, Amsterdam: North Holland.

- Shea, John (2000), “Does Parents’ Money Matter?”, *Journal of Public Economics* 77, 155-184.
- Stokey, Nancy L., and Sergio T. Rebelo (1995), “Growth Effects of Flat-Rate Taxes”, *Journal of Political Economy* 103, 519-550.
- Trostel, Philip A. (1993), “The Effect of Taxation on Human Capital”, *Journal of Political Economy* 101, 327-350.
- Ulph, David T. (1977). “On the Optimal Distribution of Income and Educational Expenditure”, *Journal of Public Economics* 8, 341-356.

# Appendix

## Second-order conditions

Substituting (1) and (3) into (4), we find the following unconstrained optimization problem - suppressing the indices  $\alpha$ :

$$\begin{aligned} \max_{e, h_1, l_2, c_1} u = & (1 + \rho) \left( \frac{c_1^{1-1/\varepsilon_c}}{1 - 1/\varepsilon_c} - \frac{(1 - h_1)^{1+1/\varepsilon_1}}{1 + 1/\varepsilon_1} \right) + (1 - t_2)l_2\phi(.) \\ & + R[(1 - t_1)(1 - e - h_1) - c_1 - (k - s)e] \\ & + (t_2 - t_1)\chi + g - \frac{l_2^{1+1/\varepsilon_2}}{1 + 1/\varepsilon_2}. \end{aligned}$$

The first-order conditions are given by:

$$\frac{\partial u}{\partial e} = (1 - t_2)l_2\phi_e(.) - R(1 - t_1 + k - s) = 0,$$

$$\frac{\partial u}{\partial h_1} = (1 + \rho)(1 - h_1)^{\frac{1}{\varepsilon_1}} - R(1 - t_1) = 0,$$

$$\frac{\partial u}{\partial l_2} = (1 - t_2)\phi(.) - l_2^{\frac{1}{\varepsilon_2}} = 0,$$

$$\frac{\partial u}{\partial c_1} = (1 + \rho)c_1^{-\frac{1}{\varepsilon_c}} - R = 0.$$

Manipulation of these first-order conditions (and using (2)) yields (5), (6), (7), and (8).

The second-order partial derivatives of the optimization are ordered in the Hessian matrix:

$$H \equiv \begin{bmatrix} (1 - t_2)l_2\phi_{ee} & 0 & (1 - t_2)\phi_e & 0 \\ 0 & -\frac{(1+\rho)}{\varepsilon_1}(1 - h_1)^{\frac{1}{\varepsilon_1}-1} & 0 & 0 \\ (1 - t_2)\phi_e & 0 & -\frac{1}{\varepsilon_2}l_2^{\frac{1}{\varepsilon_2}-1} & 0 \\ 0 & 0 & 0 & -\frac{(1+\rho)}{\varepsilon_c}c_1^{-\frac{1}{\varepsilon_c}-1} \end{bmatrix}.$$

The four leading principal minors of  $H$  are respectively:

$$(1 - t_2)l_2\phi_{ee} < 0,$$



$$\begin{aligned}
& -(1-t_2)l_2\phi_{ee}\frac{(1+\rho)}{\varepsilon_1}(1-h_1)^{\frac{1}{\varepsilon_1}-1} > 0, \\
& (1-t_2)l_2\phi_{ee}\frac{(1+\rho)}{\varepsilon_1}(1-h_1)^{\frac{1}{\varepsilon_1}-1}\frac{1}{\varepsilon_2}l_2^{\frac{1}{\varepsilon_2}-1} + \frac{(1+\rho)}{\varepsilon_1}(1-h_1)^{\frac{1}{\varepsilon_1}-1}((1-t_2)\phi_e)^2, \\
& -(1-t_2)l_2\phi_{ee}\frac{(1+\rho)}{\varepsilon_1}(1-h_1)^{\frac{1}{\varepsilon_1}-1}\frac{1}{\varepsilon_2}l_2^{\frac{1}{\varepsilon_2}-1}\frac{(1+\rho)}{\varepsilon_c}c_1^{-\frac{1}{\varepsilon_c}-1} > 0.
\end{aligned}$$

In order to guarantee a maximum, the Hessian matrix should be negative definite. Hence, the third leading principal minor must be negative. Therefore, we have:

$$-(1-t_2)l_2\phi_{ee}\frac{1}{\varepsilon_2}l_2^{1/\varepsilon_2-1} - ((1-t_2)\phi_e)^2 > 0 \Leftrightarrow -\frac{1}{\varepsilon_2}\phi_{ee}\phi(\cdot) - (\phi_e)^2 > 0,$$

where we substituted  $l_2 = ((1-t_2)\phi)^{\varepsilon_2}$  (see (8)). Using the Cobb-Douglas specification (2) of the production function of human capital, we rewrite the last inequality as (12) in the main text.

□

## Proof lemma

To prove this lemma, we write the bases of the tax on unskilled labor, capital income, and skilled labor in terms of learning (i.e. in the form of (36)) as follows:

$$\begin{aligned}
l_{1\alpha} &= 1 - h_1 - e_\alpha, \\
a_\alpha &= (1-t_1)(1-h_1) - c_1 + (1-t_1+k-s)e_\alpha, \\
\phi(\cdot)l_{2\alpha} &= \mu e_\alpha,
\end{aligned}$$

where the last two expressions follow from (1) and (11). Hence, the three tax bases (for the taxes on unskilled labor, capital income, and skilled labor) are related to learning  $e_\alpha$  in a linear fashion:

$$y_\alpha = \gamma + \pi e_\alpha, \tag{36}$$

where  $y_\alpha$  stands for the tax base and  $\gamma$  and  $\pi$  do not depend on type  $\alpha$ . Substituting this linear relationship in (16), we find:

$$\begin{aligned}\xi_y &\equiv -\frac{\left(\int_{\underline{\alpha}}^{\infty} y_\alpha b_\alpha dF - \int_{\underline{\alpha}}^{\infty} y_\alpha dF \int_{\underline{\alpha}}^{\infty} b_\alpha dF\right)}{\int_{\underline{\alpha}}^{\infty} y_\alpha dF \int_{\underline{\alpha}}^{\infty} b_\alpha dF} \\ &= -\pi \frac{\left(\int_{\underline{\alpha}}^{\infty} e_\alpha b_\alpha dF - \int_{\underline{\alpha}}^{\infty} e_\alpha dF \int_{\underline{\alpha}}^{\infty} b_\alpha dF\right)}{\int_{\underline{\alpha}}^{\infty} y_\alpha dF \int_{\underline{\alpha}}^{\infty} b_\alpha dF} = \pi \xi_e \frac{\int_{\underline{\alpha}}^{\infty} e_\alpha dF}{\int_{\underline{\alpha}}^{\infty} y_\alpha dF}.\end{aligned}$$

Note that if  $\gamma = 0$ , we have  $\int_{\underline{\alpha}}^{\infty} y_\alpha dF = \pi \int_{\underline{\alpha}}^{\infty} e_\alpha dF$  so that  $\xi_y = \xi_e$ . The lemma follows from substituting the specific linear relationships for each of the three tax bases.

□

## Generalizing the utility function

We show that the result of a zero learning wedge is robust to a general specification of the utility function that allows for income effects and cross substitution between consumption and leisure.

### Private behavior

Utility is general, and given by function  $u(c_{1\alpha}, h_{1\alpha}, c_{2\alpha}, h_{2\alpha})$  with standard properties. The private household maximizes utility with respect to the intertemporal private budget constraint:

$$\begin{aligned}&c_{2\alpha} + Rc_{1\alpha} + (1 - t_2)h_{2\alpha}\phi(\alpha; e_\alpha) + R(1 - t_1)h_{1\alpha} \\ &= (1 - t_2)\phi(\alpha; e_\alpha) + R(1 - t_1) + (t_2 - t_1)\chi + g - R(1 - t_1 + k - s)e_\alpha.\end{aligned}$$

This budget constraint implies that the normal (Marshallian) demand functions for  $c_1$ ,  $c_2$ ,  $h_1$ , and  $h_2$  are functions of the relative prices  $R$ ,  $R(1 - t_1)$ ,  $(1 - t_2)\phi(\alpha; e_\alpha)$  and an income term  $(t_2 - t_1)\chi + g - R(1 - t_1 + k - s)e_\alpha$ . The compensated (Hicksian) demand functions depend only on the three relative prices. The compensated demand function for  $y$  can thus be written as:

$$y_\alpha = \bar{y}_\alpha(R, w_1, w_2),$$

where  $y_\alpha = c_{1\alpha}^*, h_{1\alpha}^*, h_{2\alpha}^*$ ,  $w_1 \equiv (1 - t_1)$ ,  $w_2 \equiv (1 - t_2)\phi(\alpha; e_\alpha)$ . We employ asterisks to denote compensated demands. The compensated demand function implies that  $s$  affects  $y$  through the impact of learning  $e_\alpha^*$  on  $w_2$ :

$$\frac{\partial y}{\partial s} = \frac{\partial \bar{y}}{\partial w_2} (1 - t_2) \phi_e \frac{\partial e^*}{\partial s},$$

and that  $t_2$  impacts  $y$  through two channels, namely not only directly but also indirectly through its impact on learning  $e$ :

$$\frac{\partial y}{\partial t_2} = \frac{\partial \bar{y}}{\partial w_2} \left[ -\phi(.) + (1 - t_2) \phi_e \frac{\partial e^*}{\partial t_2} \right].$$

We now show that the relationship between  $\frac{\partial y}{\partial s}$  and  $\frac{\partial y}{\partial t_2}$  is independent of  $\alpha$  if the learning function is given by the Cobb Douglas form  $\phi(\alpha; e_\alpha) = \alpha e_\alpha^\beta$ . Combining the two equations above, we arrive at:

$$\frac{\frac{\partial y}{\partial s}}{\frac{\partial y}{\partial t_2}} = \frac{\frac{\partial e^*}{\partial s}}{\frac{\partial e^*}{\partial t_2} - \frac{\phi(.)}{(1-t_2)\phi_e}}. \quad (\text{A1})$$

We find expressions for  $\frac{\partial e^*}{\partial s}$  and  $\frac{\partial e^*}{\partial t_2}$  from the private first-order condition for learning:

$$(1 - t_2) l_{2\alpha}^* \phi_e(.) = R(1 - t_1 + k - s), \quad (\text{A2})$$

where  $l_{2\alpha}^* = 1 - h_{2\alpha}^*(R, w_1, w_2) = 1 - h_{2\alpha}^*(R, w_1, (1 - t_2)\phi(\alpha; e_\alpha))$ . By differentiating this private first-order condition with respect to  $s$  and  $t_2$  respectively, we find:

$$\frac{\partial e^*}{\partial s} = \frac{R}{(1 - t_2)^2 \phi_e^2 \frac{\partial h_{2\alpha}^*}{\partial w_2} - (1 - t_2) l_{2\alpha}^* \phi_{ee}} > 0, \quad (\text{A3})$$

and,

$$\frac{\partial e^*}{\partial t_2} = \frac{-l_{2\alpha}^* \phi_e + (1 - t_2) \phi(.) \phi_e \frac{\partial h_{2\alpha}^*}{\partial w_2}}{(1 - t_2)^2 \phi_e^2 \frac{\partial h_{2\alpha}^*}{\partial w_2} - (1 - t_2) l_{2\alpha}^* \phi_{ee}} < 0. \quad (\text{A4})$$

Second-order conditions require that the numerator in these equations be positive while the compensated demand derivative  $\frac{\partial h_{2\alpha}^*}{\partial w_2}$  is negative.

We subsequently substitute (A3) and (A4) into (A1) to eliminate  $\frac{\partial e^*}{\partial s}$  and  $\frac{\partial e^*}{\partial t_2}$ :

$$\frac{\frac{\partial y}{\partial s}}{\frac{\partial y}{\partial t_2}} = \frac{R}{l_{2\alpha}^* \phi_e \left[ \frac{\phi_{ee} \phi(.)}{\phi_e^2} - 1 \right]} = \frac{1 - t_2}{(1 - t_1 + k - s)} \frac{1}{\left[ \frac{\phi_{ee} \phi(.)}{\phi_e^2} - 1 \right]},$$

where the second equality follows from the first-order condition for learning (A2) by eliminating  $l_2^* \phi_e(\cdot)$ . In case of a Cobb-Douglas learning function, the term between square brackets  $\left[ \frac{\phi_{ee} \phi(\cdot)}{\phi_e^2} - 1 \right]$  amounts to  $-1/\beta$  so that:

$$\frac{\frac{\partial y}{\partial s}}{\frac{\partial y}{\partial t_2}} = -\frac{\beta(1-t_2)}{(1-t_1+k-s)}. \quad (\text{A5})$$

### Optimal lump-sum transfer

The first-order condition for maximizing social welfare with respect the lump-sum transfer is given by:

$$\begin{aligned} & \int_{\underline{\alpha}}^{\infty} \left( \frac{\Psi' \lambda_{\alpha}}{\eta} - 1 \right) dF + \Delta \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{\partial g} dF - (t_1 R + \tau r) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1}{\partial g} dF \\ & + \int_{\underline{\alpha}}^{\infty} t_2 \phi(\cdot) \frac{\partial l_2}{\partial g} dF - \tau r \int_{\underline{\alpha}}^{\infty} \frac{\partial c_1}{\partial g} dF = 0, \end{aligned}$$

where  $\lambda_{\alpha}$  is the private marginal utility of income. By defining  $b_{\alpha}$  as:

$$b_{\alpha} \equiv \frac{\Psi' \lambda_{\alpha}}{\eta} + \Delta \frac{\partial e}{\partial g} - (t_1 R + \tau r) \frac{\partial h_1}{\partial g} + t_2 \phi(\cdot) \frac{\partial l_2}{\partial g} - \tau r \frac{\partial c_1}{\partial g},$$

we can write this first-order condition as:

$$\int_{\underline{\alpha}}^{\infty} b_{\alpha} dF = 1. \quad (\text{A6})$$

### Optimal tax on skilled labor

The first-order condition for the second-period labor tax rate (i.e. the marginal tax rate on skilled labor),  $t_2$ , is given by (using Roy's identity  $\partial v_{\alpha} / \partial t_2 = -\lambda_{\alpha}(l_{2\alpha} \phi(\cdot) - \chi)$ ):

$$\begin{aligned} & \int_{\underline{\alpha}}^{\infty} \left( 1 - \frac{\Psi' \lambda_{\alpha}}{\eta} \right) (l_{2\alpha} \phi(\cdot) - \chi) dF + \Delta \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{\partial t_2} - (t_1 R + \tau r) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1}{\partial t_2} dF \\ & + \int_{\underline{\alpha}}^{\infty} \left( t_2 \phi(\cdot) \frac{\partial l_2}{\partial t_2} \right) dF - \tau r \int_{\underline{\alpha}}^{\infty} \frac{\partial c_1}{\partial t_2} dF = 0, \end{aligned}$$

where  $\Delta \equiv \frac{(1-t_1+k-s)R}{1-t_2} - (1+r)(1+k)$  represents the tax wedge on learning (i.e. the first-order effect of learning  $e$  on the government budget constraint).

Substituting the definition of  $b_\alpha$  to eliminate  $\frac{\Psi'\lambda_\alpha}{\eta}$ , we arrive at:

$$\begin{aligned}
& \int_{\underline{\alpha}}^{\infty} (1 - b) (l_{2\alpha}\phi(\cdot) - \chi) dF \\
& + \Delta \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{\partial t_2} dF + \Delta \int_{\underline{\alpha}}^{\infty} (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial e}{\partial g} dF \\
& - (t_1 R + \tau r) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1}{\partial t_2} dF - (t_1 R + \tau r) \int_{\underline{\alpha}}^{\infty} (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial h_1}{\partial g} dF \\
& + \int_{\underline{\alpha}}^{\infty} \left( t_2 \phi(\cdot) \frac{\partial l_2}{\partial t_2} \right) dF + \int_{\underline{\alpha}}^{\infty} \left( t_2 \phi(\cdot) (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial l_2}{\partial g} \right) dF \\
& - \tau r \int_{\underline{\alpha}}^{\infty} \frac{\partial c_1}{\partial t_2} dF - \tau r \int_{\underline{\alpha}}^{\infty} (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial c_1}{\partial g} dF = 0.
\end{aligned} \tag{A7}$$

We substitute the Slutsky equations:

$$\begin{aligned}
\frac{\partial e}{\partial t_2} &\equiv \frac{\partial e^*}{\partial t_2} - (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial e}{\partial g}, \\
\frac{\partial h_1}{\partial t_2} &\equiv \frac{\partial h_1^*}{\partial t_2} - (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial h_1}{\partial g}, \\
\frac{\partial l_2}{\partial t_2} &\equiv \frac{\partial l_2^*}{\partial t_2} - (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial l_2}{\partial g}, \\
\frac{\partial c_1}{\partial t_2} &\equiv \frac{\partial c_1^*}{\partial t_2} - (l_{2\alpha}\phi(\cdot) - \chi) \frac{\partial c_1}{\partial g}.
\end{aligned}$$

into (A7) to arrive at:

$$\begin{aligned}
& \int_{\underline{\alpha}}^{\infty} (1 - b_\alpha) l_{2\alpha}\phi(\cdot) dF + \Delta \int_{\underline{\alpha}}^{\infty} \frac{\partial e^*}{\partial t_2} dF - (t_1 R + \tau r) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1^*}{\partial t_2} dF \\
& + \int_{\underline{\alpha}}^{\infty} \left( t_2 \phi(\cdot) \frac{\partial l_2^*}{\partial t_2} \right) dF - \tau r \int_{\underline{\alpha}}^{\infty} \frac{\partial c_1^*}{\partial t_2} dF = 0,
\end{aligned} \tag{A8}$$

where we have employed (A6) to get rid of the term containing  $\chi$ .

### Optimal education subsidy

The first-order condition for education subsidies  $s$  amounts to (using Roy's identity  $\partial v_\alpha / \partial s = \lambda_\alpha e_\alpha R$ ):

$$R \int_{\underline{\alpha}}^{\infty} \left( \frac{\Psi' \lambda_\alpha}{\eta} - 1 \right) e_\alpha dF + \Delta \int_{\underline{\alpha}}^{\infty} \frac{\partial e}{\partial s} dF - (t_1 R + \tau r) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1}{\partial s} dF \\ + \int_{\underline{\alpha}}^{\infty} \left( t_2 \phi(.) \frac{\partial l_2}{\partial s} \right) dF - \tau r \int_{\underline{\alpha}}^{\infty} \frac{\partial c_1}{\partial t_2} dF = 0.$$

We substitute the definition of  $b_\alpha$  to eliminate  $\frac{\Psi' \lambda_\alpha}{\eta}$  and substitute the Slutsky equations:

$$\begin{aligned} \frac{\partial e}{\partial s} &\equiv \frac{\partial e^*}{\partial s} + \frac{\partial e}{\partial M} \frac{\partial M}{\partial s} = \frac{\partial e^*}{\partial s} + e_\alpha R \frac{\partial e}{\partial g}, \\ \frac{\partial h_1}{\partial s} &\equiv \frac{\partial h_1^*}{\partial s} + \frac{\partial h_1}{\partial M} \frac{\partial M}{\partial s} = \frac{\partial h_1^*}{\partial s} + e_\alpha R \frac{\partial h_1}{\partial g}, \\ \frac{\partial l_2}{\partial s} &\equiv \frac{\partial l_2^*}{\partial s} + \frac{\partial l_2}{\partial M} \frac{\partial M}{\partial s} = \frac{\partial l_2^*}{\partial s} + e_\alpha R \frac{\partial l_2}{\partial g}, \\ \frac{\partial c_1}{\partial s} &\equiv \frac{\partial c_1^*}{\partial s} + \frac{\partial c_1}{\partial M} \frac{\partial M}{\partial s} = \frac{\partial c_1^*}{\partial s} + e_\alpha R \frac{\partial c_1}{\partial g}. \end{aligned}$$

We find:

$$\begin{aligned} \int_{\underline{\alpha}}^{\infty} (b_\alpha - 1) e_\alpha dF + \frac{\Delta}{R} \int_{\underline{\alpha}}^{\infty} \frac{\partial e^*}{\partial s} dF - \left( t_1 + \frac{\tau r}{R} \right) \int_{\underline{\alpha}}^{\infty} \frac{\partial h_1^*}{\partial s} dF \\ + \int_{\underline{\alpha}}^{\infty} \left( \frac{t_2 \phi(.)}{R} \frac{\partial l_2^*}{\partial s} \right) dF - \frac{\tau r}{R} \int_{\underline{\alpha}}^{\infty} \frac{\partial c_1^*}{\partial s} dF = 0. \end{aligned} \quad (A9)$$

### Production efficiency

The first-order condition for learning implies:

$$(1 - t_2) l_{2\alpha} \phi_e e_\alpha = (1 - t_2) l_{2\alpha} \phi(.) \left( \frac{\phi_e e_\alpha}{\phi(.)} \right) = R(1 - t_1 + p - s) e_\alpha.$$

With a Cobb-Douglas learning function, we have  $\phi_e e_\alpha / \phi(.) = \beta$  so that:

$$l_{2\alpha} \phi(.) = \frac{R(1 - t_1 + p - s)}{\beta(1 - t_2)} e_\alpha \equiv \mu e_\alpha.$$

This proportional relationship between  $l_{2\alpha}\phi(\cdot)$  and  $e_\alpha$  (which does not depend on  $\alpha$ ) implies that:

$$\frac{\int_{\underline{\alpha}}^{\infty} (-b_\alpha + 1) l_{2\alpha}\phi(\alpha; e_\alpha) dF}{R \int_{\underline{\alpha}}^{\infty} (b_\alpha - 1) e_\alpha dF} = -\frac{(1 - t_1 + k - s)}{\beta(1 - t_2)}. \quad (\text{A10})$$

Substituting (A5) and (A10) into (A9), and combining the result with (A8) to solve for  $\Delta$ , we arrive at:

$$\Delta \left[ \int_{\underline{\alpha}}^{\infty} \frac{\partial e^*}{\partial t_2} + \frac{(1 - t_1 + k - s)}{\beta(1 - t_2)} \frac{\partial e^*}{\partial s} dF \right] = 0.$$

By combining (A1) and (A5) to eliminate  $\frac{\partial y}{\partial s} / \frac{\partial y}{\partial t_2}$ , we find that the term between square brackets can be written as  $\int_{\underline{\alpha}}^{\infty} \frac{\phi(\cdot)}{(1 - t_2)\phi_e} dF > 0$ . Thus, the learning wedge  $\Delta$  has to equal zero. A zero learning wedge  $\Delta \equiv \frac{(1 - t_1 + k - s)R}{1 - t_2} - (1 + r)(1 + k) = 0$  implies that expression (33) continues to hold.  $\square$

## Solution labor tax schedule with tax deductible expenses

We derive the closed for solutions (24), (25), and (26) for the specific case that education expenses are tax deductible (i.e.  $s = t_1 k$ ) in the following way. If education expenses are tax deductible, one needs to take into account of the fact that the government simultaneously changes  $s$  if  $t_1$  is affected according to  $ds = k dt_1$ . To find the optimal level of  $t_1$  (and  $s = t_1 k$ ), we combine the first-order condition for the optimal tax on unskilled labor (if education expenses are not tax-deductible) (18) and the first-order condition for optimal education subsidies (30) to arrive at:

$$\xi_e + \left( t_1 + \frac{\tau r}{R} \right) \frac{E_{1t_1}^*}{(1 - t_1)} = \frac{\Delta}{R} \frac{\varepsilon_{et_1}}{(1 - t_1)} + \frac{t_2}{R} \frac{\mu \varepsilon_{l_2 t_1}}{(1 - t_1)}, \quad (\text{A11})$$

where we have used  $\frac{\varepsilon_{es}}{(1 - s)} = \frac{\varepsilon_{et_1}}{(1 - t_1)}$  and  $\frac{\varepsilon_{l_2 s}}{(1 - s)} = \frac{\varepsilon_{l_2 t_1}}{(1 - t_1)}$  and where  $E_{1t_1}^* \equiv E_{1t_1}/(1 + k)$ .

Substitution of (A11) into (20) to eliminate the distributional characteristics  $\xi_2 = \xi_e$  (and using  $\frac{\mu \varepsilon_{l_2 t_1}}{R(1 - t_1)} = \frac{\varepsilon_{l_2 t_2}}{(1 - t_2)}$  and  $\frac{\varepsilon_{et_2}}{\mu(1 - t_2)} = \frac{\varepsilon_{et_1}}{(1 - t_1)R}\beta(1 + \varepsilon_2)$ ) yields:

$$\left( \frac{t_1 + \frac{\tau r}{R}}{1 - t_1} \right) E_{1t_1}^* = \frac{\Delta}{R(1 - t_1 + k - s)}. \quad (\text{A12})$$

Using  $\Delta \equiv \frac{(1-t_1+k-s)R}{1-t_2} - (1+r)(1+k)$ , we can write the second right-hand side of this equation as (if  $s = t_1 k$ ):

$$\frac{\Delta}{R(1-t_1)(1+k)} = \frac{1}{1-t_2} - \frac{1 + (\tau r/R)}{1-t_1}. \quad (\text{A13})$$

By substituting this expression into (A12) to eliminate  $\Delta$ , we find:

$$\left( \frac{1 + \frac{\tau r}{R}}{1-t_1} - 1 \right) E_{1t_1}^* = \frac{1}{1-t_2} - \frac{1 + \frac{\tau r}{R}}{1-t_1}.$$

Solving this expression for  $\frac{1 + \frac{\tau r}{R}}{1-t_1}$ , we arrive at:

$$\frac{1 + \frac{\tau r}{R}}{1-t_1} = \frac{\frac{1}{1-t_2} + E_{1t_1}^*}{1 + E_{1t_1}^*}.$$

Substitution of this result in (A13) to eliminate  $\frac{1 + \frac{\tau r}{R}}{1-t_1}$  yields:

$$\frac{\Delta}{R(1-t_1)(1+k)} = \frac{t_2}{1-t_2} \left[ \frac{E_{1t_1}^*}{1 + E_{1t_1}^*} \right]. \quad (\text{A14})$$

Using this expression to eliminate  $\frac{\Delta}{\mu(1-t_2)} = \frac{\Delta}{R(1-t_1)(1+k)}\beta$  (if  $s = kt_1$ ) from (20), we arrive at (24) (by using the definitions of the elasticities  $\varepsilon_{et_2} \equiv \frac{1+\varepsilon_2}{1-\beta(1+\varepsilon_2)}$  and  $\varepsilon_{l_2 t_2} \equiv \frac{\varepsilon_2}{1-\beta(1+\varepsilon_2)}$ ). Substituting (24) into (A14), we establish (26). (25) is found by using (A12), (A14), and (24).

□

## Optimal capital income tax with tax deductible expenses

If costs are deductible we have  $s = t_1 k$ . Substituting (A11) into (27), (using  $E_{1t_1}^* = \frac{\varepsilon_{(1-h_1)\tau}(1-h_1)}{(1+k)\int_{\underline{\alpha}}^{\infty} e_{\alpha} dF}$ ,  $\frac{\varepsilon_{e\tau}}{(1+k)} = \varepsilon_{et_1}$ , and  $\frac{\varepsilon_{l_2\tau}}{(1+k)} = \varepsilon_{l_2 t_1}$ ), we find:

$$\frac{\tau r}{R} \frac{\varepsilon_{c_1\tau} C_1}{\int_{\underline{\alpha}}^{\infty} e dF} = 0.$$

so that  $\tau = 0$ .

□



## Data

Table A1 - Data

Taxes (%)			
Country	$\tau^a$	$t_2^b$	$t_1^c$
Canada	38.6	45 <sup>d</sup>	36 <sup>d</sup>
Denmark	29.1	61 <sup>e</sup>	46 <sup>e</sup>
Finland	19.6	44 <sup>f</sup>	44 <sup>f</sup>
Germany	19.9	32 <sup>g</sup>	29 <sup>g</sup>
Italy	31.0	34	27
Netherlands	24.7	50	38
Sweden	30.5	31 <sup>h</sup>	31 <sup>h</sup>
United States	31.1	34.5 <sup>d</sup>	21.5 <sup>d</sup>
Education expenditures			
	Exp. in US \$ <sup>i</sup>	Proportion public (%) <sup>j</sup>	US PPP \$ <sup>k</sup>
Canada	14809	82	1.1692
Denmark	7294	95	8.33
Finland	7145	100 <sup>l</sup>	5.9
Germany	9466	79	2.003
Italy	5972	96	1594.82
Netherlands	9989	92	2.0515
Sweden	12981	98	9.6172
United States	17466	75	1
Gross wages			
	Low skilled <sup>m</sup>	High skilled <sup>n</sup>	'Return' <sup>o</sup>
Canada	23453	36588	.11
Denmark	223242	291126	.06
Finland	112808	137918	.04
Germany	30220 <sup>p</sup>	44740 <sup>p</sup>	.10
Italy (*1000)	29891	41188	.08
Netherlands	43805	57319	.06
Sweden	160089	205887	.06
United States	22688	41579	.17

Sources: OECD/Statistics Canada (1995), OECD (2000a, 2000b), IFBD (1997, 2000).

Notes:

<sup>a</sup> Average nation-wide tax on capital income in 1997. Source OECD (2000b).

<sup>b</sup> Marginal personal income tax rate on skilled labor income net of general exemptions excluding surcharges, including local taxes. Source IFBD (1997).

<sup>c</sup> Marginal personal income tax rate on unskilled labor income net of general exemptions excluding surcharges, including local taxes. Source IFBD (1997).

<sup>d</sup> For Canada and the US, figures apply to 2000 and are taken from IFBD (2000) since IBFD (1997) does not report figures for the US and Canada. We included unweighed averages of state taxes (Canada: 19%; US: 6.5%).

<sup>e</sup> Including average municipality tax (29%).

<sup>f</sup> Including municipality tax (17%).

<sup>g</sup> Based on single households without dependents.

<sup>h</sup> Including municipality tax (31%).

<sup>i</sup> Source OECD (2000a). Figures apply to 1997.

<sup>j</sup> Source OECD (2000a). Figures apply to 1997.

<sup>k</sup> PPP deflator by OECD is used to transform education expenditures in local currency.

<sup>l</sup> No figures for the share of education expenditures were available in OECD (2000a). We therefore employed the value of public education expenditures from OECD (1996).

<sup>m</sup> Average yearly gross wages of male workers with less than 15 years of education. Source OECD/Statistics Canada (1995).

<sup>n</sup> Average yearly gross wages of male workers with more than 15 years of education. Source OECD/Statistics Canada (1995).

<sup>o</sup> Approximation of the average return to higher education measured as the percentage increase in wages (assuming that education takes 5 years).

<sup>p</sup> For Germany wages are reported only in after-tax terms (unskilled DM 24950; skilled DM 35180). We computed gross wages by linearly approximating the tax schedule taking into account the general income exemption for single households.